

> 0 <
OI IO IntelliGenetics
> 0 <

Quest - Quick User-directed Expression Search Tool
Release 5.4

-- Outline of search "seq4" --

Selected search type is key against sequence data banks or files.

Selected scope is Sequence.

Selected sequence key from "pappu403.key":

seq4 (NA) ID seq4 NA preliminary pattern

1 either-or
2 aacgttcg
2 cgaacgtt

Selected files:

File : Immun_ge.seq

-- Output Parameters --

Format Options: File Options:

| | | | |
|-----------------------------|-------|----------------------|-----|
| Nucleic acid code matching | Exact | Indirect file | No |
| Find non-matching hits only | No | Sequence or key file | No |
| Report key used | Yes | List of hits | Yes |
| Note position of hit | Yes | Hit display | Yes |
| Display full annotations | Yes | Name and annotations | Yes |
| Sequence context | 50 | | |

-- Run Parameters --

Run mode Batch
Time to start comparison now
Notify at end of run No

1 match found in sequence:

af058689 : TOIG of: af058689 check: 5450 from: 1 to: 29831
(from "Immun_ge.seq")

TOIG of: af058689 check: 5450 from: 1 to: 29831

LOCUS AF058689 29831 bp DNA BCT 26-MAY-2000
DEFINITION Neisseria meningitidis strain 22491, genomic sequence.

ACCESSION AF058689

VERSION AF058689.2 GI:6274529

KEYWORDS

SOURCE

ORGANISM

Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.

1 (bases 2883 to 7798)

Brieske.N., Schenker.M., Schnibbe.T., Quentin-Millet,M.J. and

Achtman.M.

Human antibody responses to A and C capsular polysaccharides, IgM

protease and transferrin-binding protein complex stimulated by

infection with Neisseria meningitidis of subgroup IV-1 or ET-37

complex

Vaccine 17 (7-8), 731-744 (1999)

MEDLINE 99165001

PUBMED 10067678

REFERENCE 2 (bases 1 to 29831)

Lin,Z.B., Schenker.M., Zhu,P. and Achtman.M.

Frequent interspecific genetic exchange between commensal

neisseriae and neisseria meningitidis

Mol. Microbiol. 36 (5), 1049-1058 (2000)

JOURNAL MEDLINE 20305048

REFERENCE 3 (bases 2883 to 7798)

Schenker,M. and Achtman,M.

Direct Submission

Submitted (09-APR-1998) Max-Planck-Institut fuer molekulare

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

Genetik, Ihnestr. 73, Berlin 14195, Germany

4 (bases 1 to 29831)

Schenker,M., Morelli,G. and Achtman,M.

Direct Submission

Submitted (08-NOV-1999) Trautner, AG Achtman, Max-Planck-Institut

fuer molekulare Genetik, Ihnestr. 73, Berlin 14195, Germany

Sequence update by submitter

On Nov 8, 1999 this sequence version replaced gi:3047409.

Location/Qualifiers

1..29831

/organism="Neisseria meningitidis"

/strain="22491"

/serotype="4,21"

/db_xref="taxon:487"

/map="between potF and pdh_tyrC"

/tissue_type="cerebrospinal fluid"

/country="Gambia"

/note="serosubtype Pl.7b,13a, immunotype L9

isolated from patient with meningitis; formerly designated

C751

serogroup: A; subgroup: IV-1"

<1..756

/gene="potF"

<1..756

/gene="potF"

/codon_start=1

/transl_table=11

/product="putative periplasmatic putrescin transporter

PotF"

/protein_id="AAF06678.1"

/db_xref="GI:6274530"

/translation="LKMLETADPGNQYAVYFSGVNTAIATAGKELLGKLPENGWD

LLFKPEYTHKLGKSCGIALWDTPEMPILNLVGLKDPKGSNPDLKAAAEVLKSIKRPD

VKRFSPSIIDELAKGICLAAGNGDLNLAKESEVKNNVGLVLPKMGFIWSM

LIPADAKNVANAKHYINYITLDPEIAKNGIATVTFAPASKPAREKMPALVNTKSIFFN

EQDMKDFVMPQNMSTDAKKLSVLSLQKIKVGTN"

828..984

/note="Correia element CE_tol"

/rpt_type="dispersed

1226..2882

/note="neisserial intergenic mosaic element (NIME)"

/rpt_type="direct

1226..1245

/rpt_family="dRS3"

/rpt_type="direct

1246..1279

/note="rs-element RS14"

1280..1299

/rpt_family="dRS3"

/rpt_type="direct

1300..1355

/note="rs-element RS26"

1356..1375

/rpt_family="dRS3"

/rpt_type="direct

1379..1477

/note="rs-element RS25"

/rpt_family="dRS3"

/rpt_type="direct

1498..1551

/note="rs-element RS22"

1552..1571

/rpt_family="dRS3"

/rpt_type="direct

1572..1625

/note="rs-element RS29"

1626..1645

/rpt_family="dRS3"

/rpt_type="direct

1646..1704

/note="rs-element RS14"

1705..1724

```

/rpt_family="dRS3"
/rpt_type-direct
1725..1780
/notes="rs-element RS26"
1781..1800
/rpt_family="dRS3"
/rpt_type-direct
1801..1856
/notes="rs-element RS30"
1857..1876
/rpt_family="dRS3"
/rpt_type-direct
1877..1928
/notes="rs-element RS23"
1929..1948
/rpt_family="dRS3"
/rpt_type-direct
1949..2002
/notes="rs-element RS22"
2003..2022
/rpt_family="dRS3"
/rpt_type-direct
2023..2078
/notes="rs-element RS26"
2079..2098
/rpt_family="dRS3"
/rpt_type-direct
2099..2133
/notes="rs-element RS14"
2134..2153
/rpt_family="dRS3"
/rpt_type-direct
2154..2209
/notes="rs-element RS26"
2210..2229
/rpt_family="dRS3"
/rpt_type-direct
2230..2284
/notes="rs-element R24"
2285..2304
/rpt_family="dRS3"
/rpt_type-direct
2305..2360
/notes="rs-element RS26"
2361..2380
/rpt_family="dRS3"
/rpt_type-direct
2381..2500
/notes="rs-element RS28"
2501..2520
/rpt_family="dRS3"
/rpt_type-direct
2521..2554
/notes="rs-element RS14"
2555..2574
/rpt_family="dRS3"
/rpt_type-direct
2575..2635
/notes="rs-element RS15"
2636..2655
/rpt_family="dRS3"
/rpt_type-direct
2656..2755
/notes="rs-element RS27"
2756..2775
/rpt_family="dRS3"
/rpt_type-direct
2776..2885
/notes="rs-element RS28"
complement(2883..5615)
/genes="tbpA"
complement(2883..5615)
/genes="tbpA"

```

```

/codon_start=1
/transl_table=11
/product="transferrin binding protein A precursor"
/protein_id="AAC13726.1"
/db_xref="GI:3047411"
/transl_table="MQOHLFRNLILSLMTALPVVAENVQAOQAEKQDITQVKA
KKQTRNRDNEVTGLKLVKTADTILSKQVLDIRDLTRYDPCIADVVEGRGASSYSIR
GMDKNRVSLTVGDLAQIQSYTAQAALGTAGSSGAINIEIYENVKAVEISKGNSV
EOGSGALAGSVAFQTKTADDDVIGEGROWGQOSKTASGKNRGLTOSTALAGRIGAAEA
LLIHTGRHAGEIRAHEDAGRGVOSFNRLVPVDETHPYANFIVEECESGNYETCKSP
KDVVCKDERQTVSTRDYTGPNRFLADPLSYESRSWLFPRGFRPNKRHHYIGGLEHT
QQTFDTRMTVPAPLTKAVEFENRKYIPINGNKIYAGDHRYGGLFTNGENALVGAET
GTGVYDETHTKRYGLEIYVYTNADKTDWDYARLSYDRQIGIDNHFQTHCSADGS
DKYCRPSADPFYSYKSKTPKTPANPGKQDPYVWSIGRNVVVGQICRLGNNTYTD
LRHQDYVYOHANRAYSSKTPKTPANPGKQDPYVWSIGRNVVVGQICRLGNNTYTD
CTPRSLNGKSYAAVRDNRVLRGWADVAGLRVDRYRSHDSDGSGVGTHTLSWVAG
IVLKPTDMLDLYTSTGFRPLPSFAENYGNRAGVQSKAVKIDPKSFNKKEAGLVFKGD
FGNLEASWFNAYRDLIVRPEAQIQNGKETAKGDPAYLNAQSRKINAGIIVKIDWN
GVWDKLPBGWYSTFAFVNRVHVRDIKKRADRTDIQSHLFDAIQPSRYGVYDQPEGK
WGVNGMLTYSKAKEITELLSGSRALLNGNSRNTKATARRTPWYIVDYVSGYTIKKHFT
LRAGVYNLLNRYVVTWENVROTAGGAVNQHKNVVYNYAAPGRNYTFSLEMKE"
complement(2886..5543)
/genes="tbpA"
/product="transferrin binding protein A"
complement(5544..5615)
/genes="tbpA"
complement(5702..7798)
/genes="tbpB"
complement(5702..7798)
/genes="tbpB"
/codon_start=1
/transl_table=11
/product="transferrin binding protein B precursor"
/protein_id="AAC13725.1"
/db_xref="GI:3047410"
/transl_table="MNNPLVQAAMVLPVLLSACLGGGSGFDLSDVDTEAPRAPKY
QDVSEKPAQKDDGYGFEAMLRNRVYQANPKDEIKLSSENDWEQTDNGDKNPS
KOKNIINALTNGNGASLQDSQENQGISKVTGHHNFQYVWSGFFYKQIGSTFERKDS
SITAARSGPDGYIFYKQKDP SKLPGVGEVYKGTWDFLTDVTSOKFTDLGNTSTRP
GDRYSAFSGELDYIVNKDSKDEHVGWGLTETITVDFEKTLTSLGLIKNNRVNNDND
KHTQVYSIDATLRGRNFSKAEATDKPKNDGETKEHPFVSDSSLSLGGFFGPKGBEL
GFRLSDQKVAVGSAAKTQDKERNAGAVSGGAGAAASAGTSSNSKLTITLDVAI
ELTLDNKKIKNLDNFNAALVVDGIMIPLLPEASEGKNQNOGTNGTAFTRKFNH
TPKSDKEDTQAGTAENGNAASNTAGDANGTKTYEVEVCCSNLNLKYGLMTRKNSK
SAMQGESSSQADAKTEQVQSFMFLQGETDEKEIPNDQNVVYRGSWYGIANGTWSK
GNASDKEGGNRADFTVNGTKKINGTLTADNROAAFTIIVGDIEGNGFCGTAKTADSG
FDLDQSNTRTPKAYITNAKVQGGFYGPRAEELGGWFAYSDDKQTKNATDASNGNSA
SSATVVFAGAKRQKPVQ"
complement(5705..7738)
/genes="tbpB"
/product="transferrin binding protein B"
complement(7739..7798)
/genes="tbpB"
7905..9896
/notes="neisserial intergenic mosaic element (NIME)"
/rpt_type-direct
7905..7924
/rpt_family="dRS3"
/rpt_type-direct
7925..7978
/notes="rs-element RS14"
7979..7998
/rpt_family="dRS3"
/rpt_type-direct
7999..8050
/notes="rs-element RS23"
8051..8070
/rpt_family="dRS3"
/rpt_type-direct
8071..8126
/notes="rs-element RS24"
8127..8146
/rpt_family="dRS3"

```


repeat_region /rpt_type=direct
8147. .8201
/note="rs-element RS24"
repeat_region 8202. .8221
/rpt_family="dRS3"
/rpt_type=direct
8222. .8282
/note="rs-element RS15"
repeat_region 8283. .8302
/rpt_family="dRS3"
/rpt_type=direct
8303. .8361
/note="rs-element RS14"
repeat_region 8362. .8381
/rpt_family="dRS3"
/rpt_type=direct
8382. .8437
/note="rs-element RS26"
repeat_region 8438. .8457
/rpt_family="dRS3"
/rpt_type=direct
8458. .8516
/note="rs-element RS14"
repeat_region 8517. .8536
/rpt_family="dRS3"
/rpt_type=direct
8537. .8592
/note="rs-element RS26"
repeat_region 8593. .8612
/rpt_family="dRS3"
/rpt_type=direct
8613. .8671
/note="rs-element RS14"
repeat_region 8672. .8691
/rpt_family="dRS3"
/rpt_type=direct
8692. .8743
/note="rs-element RS23"
repeat_region 8744. .8763
/rpt_family="dRS3"
/rpt_type=direct
8764. .8817
/note="rs-element RS22"
repeat_region 8818. .8837
/rpt_family="dRS3"
/rpt_type=direct
8838. .8894
/note="rs-element RS26"
repeat_region 8895. .8914
/rpt_family="dRS3"
/rpt_type=direct
8915. .8969
/note="rs-element RS24"
repeat_region 8970. .8989
/rpt_family="dRS3"
/rpt_type=direct
8990. .9045
/note="rs-element RS26"
repeat_region 9046. .9065
/rpt_family="dRS3"
/rpt_type=direct
9066. .9167
/note="rs-element RS25"
repeat_region 9168. .9187
/rpt_family="dRS3"
/rpt_type=direct
9188. .9239
/note="rs-element RS23"
repeat_region 9240. .9259
/rpt_family="dRS3"
/rpt_type=direct
9260. .9315
/note="rs-element RS26"

repeat_region 9316. .9335
/rpt_family="dRS3"
/rpt_type=direct
9336. .9370
/note="rs-element RS14"
repeat_region 9371. .9390
/rpt_family="dRS3"
/rpt_type=direct
9391. .9446
/note="rs-element RS26"
repeat_region 9447. .9466
/rpt_family="dRS3"
/rpt_type=direct
9467. .9493
/note="rs-element RS25"
repeat_region 9494. .9513
/rpt_family="dRS3"
/rpt_type=direct
9514. .9569
/note="rs-element RS26"
repeat_region 9570. .9589
/rpt_family="dRS3"
/rpt_type=direct
9590. .9648
/note="rs-element RS14"
repeat_region 9649. .9668
/rpt_family="dRS3"
/rpt_type=direct
9669. .9724
/note="rs-element RS24"
repeat_region 9725. .9744
/rpt_family="dRS3"
/rpt_type=direct
9745. .9772
/note="rs-element RS27"
repeat_region 9773. .9792
/rpt_family="dRS3"
/rpt_type=direct
9793. .9848
/note="rs-element RS26"
repeat_region 9849. .9868
/rpt_family="dRS3"
/rpt_type=direct
9869. .9896
/note="rs-element RS27"
complement(9917. .10729)
/gene="muri"
CDS complement(9917. .10729)
/gene="muri"
/codon_start=1
/transl_table=11
/product="putative glutamate racemase Muri"
/protein_id="AAF06679.1"
/db_xref="GI:6274531"
translation="MENIGRQRPVGVDSGIGGLTNVRLMERLPMENIIFYGDTARV
PYGTSKATIEFNSMQIVDFLEHDVKAMVIACNTAAVAGQKIRQKTNMPVDVTS
AGAKAALATRNKNGIGIATNTVNSNAYARATHRNPDTLVRTQAPLVLPLVEEGW
LEHEVTRLIVCEYLKPELLADGIDTLVLGCTHFPLLKPLIGRAGNVALVDSAITTARE
TARVLAQEGELNNDNNPDYFRFYSDIPLAFRTIGERFLGRTMEQIEMWSLG"
11263. .12639
/gene="amic"
11263. .12639
/gene="amic"
/codon_start=1
/transl_table=11
/product="putative N-acetylauramoyl-L-alanine
amidase-precursor Amic"
/protein_id="AAF06680.1"
/db_xref="GI:6274532"
translation="MAATGRRIYPARRHHNRNIDTRRRRQKMPADRPYFTRKPAAMT
KLTRRQIIRRTAGTLFALSPASAVKTVRAPOFTAVRIMPSTYTRLTLESTALKY
QHFTLDNPGRLVVDIQNANTVILHGLSQVKMADDPFIRNIRAGONTPTDTRLVLDLK
QPSYAQVESLPPVGGFKNRLVVDLYPQGMADDPMMALLNGSLNKLTRGSPSEADLAQN


```

CDS
/ gene="pheA"
complement(20083..21210)
/ gene="pheA"
/ codon_start=1
/ transl_table=11
/ product="putative chorismate mutase/prephenate
dehydratase PheA"
/ protein_id="AAF06690.1"
/ db_xref="GI:6274542"
/ translation="MLECTANHRSGEIMSQTIDELLPHRNAIDTIDAEILRLNERRA
QHAAIGELKGTGAVRPEREVARLRRIQDLNKGPLDPEVSARLREVMSECLAVERP
LTIAYLPGQGTQQAALKHFGHAHTWACPTIDNCFKQVETROADYLVAPEVNSTEG
SVGRILDLAVTALQACGEVLRIRHNLLRKNNGSTEGIAKFVSHAQALAQCNWLGR
HLPNARTAVSNAEARLVAESDDGTAAATAGRTAAEIIYGLDMVAECIEDFPNNTR
FLVMGHETGASGDKTSLAVAPNRGAVASLLQPLTESGISMFKFESRPSKSVLIWE
YLFFDIEGHRDAQIQTALERLGERASRVKVGISYPTAVL"
/ gene="bcr"
complement(21249..22481)
/ gene="bcr"
complement(21249..22481)
/ gene="bcr"
/ codon_start=1
/ transl_table=11
/ product="putative bicyclomycin resistance protein homolog
Bcr"
/ protein_id="AAF06691.1"
/ db_xref="GI:6274543"
/ translation="MPSAHYPEMSEKMLMAMLVMTLMPFSIDAYLPAIPEMAQSLN
ADVRIEQLSLFMFGTAFGVGVGSVDIKGRKPVALTGLAVYCLAVAAIVFASSAE
OLLNLRVQAGFAGMTVIVGAMVRDYSGRKAOMEALIGIILMVVPLVAPMGALL
QGLGQWAIFVELAAYSLVLLGLVOYFLPKPAVGKIGRDVGLVAGRKRVLKTRAA
MGLFFQAFSGMFAFUTESSFYQOOLYHVTPHQAFAFALNIITMMFFNRTAWRL
KTGVHPQSTILMGIVYQFAANLSQAAVLFFGLPPFWLLVACVMFSVGTQGLVGANTQ
ACFMSYFKEEGGSANAVLGVFQSLIGAGVGMAATLHDGSAVTMAATWTASTSGIAL
LWLCSHRAWKENGQSEYL"
complement(22615..22768)
/ note="Correia element CE_to3"
/ rpt_type=dispersed
22777..22796
/ rpt_family="drs3"
/ rpt_type=direct
22777..22945
/ note="neisserial intergenic mosaic element (NIME)"
/ rpt_type=direct
22797..22852
/ note="rs-element RS26"
22853..22872
/ rpt_family="drs3"
/ rpt_type=direct
22873..22925
/ note="rs-element RS31"
22926..22945
/ rpt_family="drs3"
/ rpt_type=direct
complement(23000..23767)
/ gene="tou6"
complement(23000..23767)
/ gene="tou6"
/ note="hypothetical protein; ttpAB-opaA region"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="Tou6"
/ protein_id="AAF06692.1"
/ db_xref="GI:6274544"
/ translation="MWGVRLIIAALPALPACRRHVKGKIPREWKPELLIVSFVNYVLT
LIQFVGLKTYTSAASAVIVGLEPLLMVFVGHFFNDKRAYHWICGAFAFVALLIMA
GGAEGGVGFVGCCLLVLAGAGCAAMRPTQRLIARIGAPFTSVSIAAASLMCLPF
SIALAQSTVTDWSVGMVLSLLYLVGVCSWYAYWLWNGMRVPANVSGLLISLEPVVG
VLLAVLILGEHLSPVSVLGVFVIAATLVAGRLSHQK"
23890..24043
/ note="Correia element CE_to4"
/ rpt_type=dispersed
24062..25439

```

```

repeat_region
/ note="neisserial intergenic mosaic element (NIME)"
/ rpt_type=direct
24062..24081
/ rpt_family="drs3"
/ rpt_type=direct
24082..24137
/ note="rs-element RS17"
24138..24157
/ rpt_family="drs3"
/ rpt_type=direct
24158..24213
/ note="rs-element RS18"
24214..24233
/ rpt_family="drs3"
/ rpt_type=direct
24234..24303
/ note="rs-element RS16"
24304..24323
/ rpt_family="drs3"
/ rpt_type=direct
24324..24379
/ note="rs-element RS13"
24380..24399
/ rpt_family="drs3"
/ rpt_type=direct
24400..24469
/ note="rs-element RS16"
24470..24489
/ rpt_family="drs3"
/ rpt_type=direct
24490..24545
/ note="rs-element RS13"
24546..24565
/ rpt_family="drs3"
/ rpt_type=direct
24566..24671
/ note="rs-element RS14"
24672..24691
/ rpt_family="drs3"
/ rpt_type=direct
24692..24747
/ note="rs-element RS17"
24748..24767
/ rpt_family="drs3"
/ rpt_type=direct
24794..24948
/ note="Correia-element CE_to5"
/ rpt_type=dispersed
24969..24988
/ rpt_family="drs3"
/ rpt_type=direct
24989..25044
/ note="rs-element RS17"
25045..25064
/ rpt_family="drs3"
/ rpt_type=direct
25065..25120
/ note="rs-element RS18"
25121..25140
/ rpt_family="drs3"
/ rpt_type=direct
25141..25210
/ note="rs-element RS16"
25211..25225
/ rpt_family="drs3"
/ rpt_type=direct
25226..25276
/ note="rs-element RS14"
25277..25296
/ rpt_family="drs3"
/ rpt_type=direct
25297..25352
/ note="rs-element RS17"

```



```
ACCESSION      AF072783
VERSION        AF072783.1  GI:3328088.
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus

REFERENCE
AUTHORS        Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE          1 (bases 1 to 270)
AUTHORS        Krishnan,M.R. and Marion,T.N.
TITLE          Comparison of the frequencies of arginines in heavy-chain CDR3 of
                antibodies expressed in the primary B cell repertoires of
                autoimmune-prone and normal mice
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 270)
AUTHORS        Krishnan,M.R. and Marion,T.N.
TITLE          Direct Submission
JOURNAL        Submitted (16-JUN-1998) Microbiology and Immunology, The University
                of Tennessee, Memphis, 858 Madison Ave., Memphis, TN 38163, USA
FEATURES       1. .270
                Location/Qualifiers
                /organism="Mus musculus"
                /strain="(NZB x NZW)F1"
                /db_xref="taxon:10090"
                /cell_type="B cell hybridoma"
                /cell_line="375.57"
                /rearranged
                /note="generated from LPS-stimulated spleen cells"
                <1..>270
                /codon_start=2
                /product="monoclonal anti-DNA IgM kappa-chain variable
                region"
                /protein_id="AAC26776.1"
                /db_xref="GI:3328089"
                /translation="SISCRSSQSIHSGNTYLEWYLOKPGQSPKLLIYKVNRFSGV
                PDRFSGSGSTDTLKISRVEAEDLGYVCQQSHPVTEGGGKTL"
BASE COUNT    72 a 63 c 66 g 68 t 1 others
ORIGIN

AF072783 Length: 270 February 15, 2002 13:34 Type: N Check: 2352
Found using 'seq4' (pappu403.key)

...

195  AGGCTGAGGATCTGGAGTTTATTACTGCTNTCAAGTTACACATGTTCTCGAAGCTCG
      |-----|
      |-----|
      245 252
      247 254

255  GTGAGGCACCAAGCT

-----
17 matches found in sequence:
ap001517 ; TOIG of: ap001517 check: 8703 from: 1 to: 294250
(from "immun-ge.seq")
TOIG of: ap001517 check: 8703 from: 1 to: 294250

LOCUS      AP001517 294250 bp DNA BCT 10-JAN-2001
DEFINITION Bacillus halodurans genomic DNA, section 11/14.
ACCESSION  AP001517 BA000004
KEYWORDS
SOURCE     Bacillus halodurans DNA.
ORGANISM   Bacillus halodurans
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  1 (sites)
AUTHORS    Takami,H. and Horikoshi,K.
TITLE      Reidentification of facultatively alkaliphilic Bacillus sp. C-125
            to Bacillus halodurans
JOURNAL    Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
REFERENCE  2 (sites)
AUTHORS    Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,
            Horikoshi,K.
            Location/Qualifiers
            1. .294250
            /organism="Bacillus halodurans"
            /db_xref="taxon:86665"
```

```
TITLE      Sasaki,R., Hirama,C., Fuji,F. and Masui,N.
            Genetic analysis of the chromosome of alkaliphilic Bacillus
            halodurans C-125
JOURNAL    Extremophiles 3 (3), 227-233 (1999)
REFERENCE  99411980
            3 (sites)
AUTHORS    Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F.,
            Nakamura,Y. and Inoue,A.
TITLE      An improved physical and genetic map of the genome of alkaliphilic
            Bacillus sp. C-125
JOURNAL    Extremophiles 3 (1), 21-28 (1999)
REFERENCE  99184645
            4 (sites)
AUTHORS    Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.
TITLE      Replication origin region of the chromosome of alkaliphilic
            Bacillus halodurans C-125
JOURNAL    Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
REFERENCE  99356711
            5 (sites)
AUTHORS    Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and
            Horikoshi,K.
TITLE      Sequence analysis of a 32-kb region including the major ribosomal
            protein gene clusters from alkaliphilic Bacillus sp. strain C-125
JOURNAL    Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
REFERENCE  99209008
            6 (sites)
AUTHORS    Takami,H.
TITLE      Genome analysis of facultatively alkaliphilic Bacillus halodurans
            C-125
JOURNAL    (in) Extremophiles in deep-sea environments (Ed.):
            HORIKOSHI, K. TSUJII;
            : 249-284; Springer-Verlag (1999)
            7 (sites)
AUTHORS    Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
            Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.
TITLE      Sequencing of three lambda clones from the genome of alkaliphilic
            Bacillus sp. strain C-125
JOURNAL    Extremophiles 3 (1), 29-34 (1999)
REFERENCE  99184646
            8 (sites)
AUTHORS    Takami,H. and Horikoshi,K.
TITLE      Analysis of the genome of an alkaliphilic Bacillus strain from an
            industrial point of view
JOURNAL    Extremophiles 4 (2), 99-108 (2000)
REFERENCE  20263314
            9 (sites)
AUTHORS    Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,
            Hirama,C., Fuji,F. and Takami,H.
TITLE      Characterization and comparative study of the rrn operons of
            alkaliphilic Bacillus halodurans C-125
JOURNAL    Extremophiles 4 (4), 209-214 (2000)
REFERENCE  20426005
            10 (sites)
AUTHORS    Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,
            Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and
            Horikoshi,K.
TITLE      Complete genome sequence of the alkaliphilic bacterium Bacillus
            halodurans and genomic sequence comparison with Bacillus subtilis
            Nucleic Acids Res. 28 (21), 4317-4331 (2000)
JOURNAL    20512582
            11 (bases 1 to 294250)
AUTHORS    Takami,H. and Takaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and
            Technology Center, Deep-sea Microorganisms Research Group; 2-15
            Natsushima, Yokosuka, Kanagawa 237-0061, Japan
            (E-mail:takami@jamstec.go.jp)
            URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearchch.html.
            Tel.:81-468-67-3895, Fax:81-468-66-6364)
FEATURES       Location/Qualifiers
                1. .294250
                /organism="Bacillus halodurans"
                /db_xref="taxon:86665"
```

gene /note="alkaliphile"
236. .823
/gene="BH2879"
CDS 236. .823
/gene="BH2879"
/codon_start=1
/transl_table=11
/product="spore cortex-lytic enzyme"
/protein_id="BAB06598.1"
/db_xref="GI:10175501"
/translation="MMKFEVLLFSCIIILSFTLTNKAESASHKHQVQSGDITLYLLSEQV
GVPMEAKRINERSNNTIYRGEOLTIPARVTASERDLIARLVHAEAGEPYPAGKVAVA
VVLNRVDHPSFPDQNTDVINEVSGYVAFSPVONGRIQRPANDEAKRANDEALQFRG
QGNGSLYFNPKTATSPYVSTROQTIVIGNHIFAK"
955. .1605
/gene="BH2880"
CDS 955. .1605
/gene="BH2880"
/note="BH2880"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06599.1"
/db_xref="GI:10175502"
/translation="MRRISRTSSRTGYRATYVYNPLGTSDSVRKVEPPPIETANET
AKPSNNVLSYERKQYLKELKQVFKSPYRHEKELYHTLQAIGNKEELISQTCITLIC
KYNELKSIRKLDQFQTSYMDIOALHIQHQKEALEAGLLNDNGTLLNSDVFRQ
LSTSDQDTMAVLKRFKSLLLKQYHTISQLOVPNSHOMSPYDPPGEPLKGFIEEKG"
1694. .2977
/gene="BH2881"
CDS 1694. .2977
/gene="BH2881"
/note="BH2881"
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/protein_id="BAB06600.1"
/db_xref="GI:10175503"
/translation="MSEEHQESHSDNLDLMNHTLTNDHLQLLKLIGEMAALKRML
FLKQTVRMDLVGPGGDLVIEGDALAFSONVANVLGGKVKHHPFATATVWGAEN
LKDIVSARAESYAKGALPTIRHSHITDRLARRDFSINAMATHLHPASYGQLVDPPH
GRHDLTNGLRILIHQSQFIDDPTRLLRGVRFRNFYEQKTANLALATQALPTNAL
ANYSPIRVELKILCHETDPVSSEKLEDLHWQALLGLTFSSSSATHLSRLOEON
GEPLHFOATATVGFEDNMKASLVPATVAMEQRFQNIEDIQKRLTNMTRESTDYL
HKQIQVPEPLRFYALSSGEEKQVLDLYLHQRKQLPLLTGHDLMELGMKPSPLFK
ECLLHECEQLKGTIENKODALQFAREFFNHKQPL"
complement(3074. .4309)
/gene="BH2882"
CDS complement(3074. .4309)
/gene="BH2882"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier protein) synthase"
/protein_id="BAB06601.1"
/db_xref="GI:10175504"
/translation="MRRVVVTGIGAVTPLGNTABETWEKAIAGQSVGLLTRVEAAD
FPMKVAEVEFPTTHIDRKEARKMDRFTQFAVASALMAKADADLEITEDIAPRGV
WICSGGGMETYNQRFMEIKGYRRVSPFVPMITPDMAAGQVSIITGARGINSCSV
TACAGANSITGDAFKVIQRGDADVMITGGAEAPITNNVAVGFSMAKAISTDSPNTAS
RPFDAHRDGVGMGEGGILLILESGAERKAKIYAEIYGVGATGDAYHVTPQAPERGE
GAARMKQATADAGLTPEIDQYVNAHGTSTPYNDKYETLAAKQVFGDHYNNLAISSTK
SMTGCHLLAGAGVAIFSYKAEIEGIIPPTINYTEPDDPCDLDIYVPNEARKQGVNAV
SNSLGFGGHNATLVFKKYQ"
complement(4347. .5285)
/gene="BH2883"
CDS complement(4347. .5285)
/gene="BH2883"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier protein) synthase"
/protein_id="BAB06602.1"
/db_xref="GI:10175505"
/translation="MTRAGILIGISYVGEHVHTNKDFEKRIDTSDIEWIRTRTGIEERR

FAPDNVDSMAYEASVKALKAAGVSAREIEDLILVATVTPDMPPPTVSTLVQORLGAK
KAAAMDISACAGSEFIYLATQQCFIENGQYRHLVIGVKEKLSKITDMTRNTAVLFGD
GAGAAVMGAVSEDRGILAYELGADGEGAMHINQKGEYIQMNGREVFKFAVRQMGESAL
SVLEKAGLSKEDVDFLIPHQANIRIMEASRRERLELPVEKMTSTTVKKYGTNTSSASIPMA
MVDELKDGKIKEDGDLVLVGFAGLVWGSLALRWGR"
5508. .5807
/gene="BH2884"
CDS 5508. .5807
/gene="BH2884"
/note="BH2884"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06603.1"
/db_xref="GI:10175506"
/translation="MSNQERDFFTDLMFGRPPSTDAEENPOESTSGSTVEETKTE
GETKQEETNAAPALPLTFEQMEHIFRLAQSLGPAKLSLSPVVKTIQQLISQQKND"
complement(5903. .6157)
/gene="BH2885"
CDS complement(5903. .6157)
/gene="BH2885"
/note="BH2885"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06604.1"
/db_xref="GI:10175507"
/translation="MNVYTHGPIYRQYGYIGTGERFYPVLPFLAGLAVGPLLFGGKP
YGGYPAGPSYGNPFGPNYGPNGFPMYGPSPQPPFGYYK"
complement(6226. .6555)
/gene="BH2886"
CDS complement(6226. .6555)
/gene="BH2886"
/note="BH2886"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06605.1"
/db_xref="GI:10175508"
/translation="MKNKDALRSFVKVKKGKQOLKDKRFEEGIQTLTPYTELFROTD
VAPQVFSVYAIQLRGTGEFGLRTVEDIKGMELKTEAEVKAVEKLEGFLHDVLAQL
ASSDKQL"
complement(6636. .7421)
/gene="BH2887"
CDS complement(6636. .7421)
/gene="BH2887"
/note="BH2887"
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/protein_id="BAB06606.1"
/db_xref="GI:10175509"
/translation="MGI GSSNRQTDQNNAKLYDERHFRFVSAYGCDLLIOWLAPKEGCV
LDLGCCTGDLTEQIHQLSRVIGVDVSESMIEQAQKGFPHLDQVAAATDLSSEKFD
WISFNVLHWIKQAEALTVIYSLKPGGRFVAFEGGKGNVEITVNTLADTMSQADSR
AFSKEELUPWFFPSIQYTSLMEQGVFRVTHALPHFNRTLLEDGGGIRHFLAMFAESVLE
GKSTERAETVREVAEVDKALDFDGTNNWADYKRIRYVVGKQL"
complement(7607. .7843)
/gene="BH2888"
CDS complement(7607. .7843)
/gene="BH2888"
/note="BH2888"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06607.1"
/db_xref="GI:10175510"
/translation="MYEHQNVNVTQVPEPNRNMQIAMKHLPEGKAFLDKKGTEL
SMEEMQPMQLLNNVNAEAYELGWEDAQQSQNDG"
complement(7879. .8751)
/gene="med"
CDS complement(7879. .8751)
/gene="med"

/note="BH2889"
/codon_start=1
/transl_table=11
/product="transcriptional activator of comk gene"
/protein_id="BAB06608.1"
/db_xref="GI:10175511"
/translation="MYGLLVEDITIDQGNRKAYEGLLIHNSLNDVVVLEEGVNSEQ
KAHRRITKELVDGCVNLIFGHHAFAYFSTIHQYPDVHFVFNFGVKENITSLHFE
GYAMGYFGMAASSETHKVGVIAAFPMQVEGVDGAKYMNESEAFVRVXGEMTD
ADKALEFLOKEQVDVEYPAGDGYHPVVEAIKDQGFALGYGVDQADLGGSTILT
SFVQVDDLIYLVAKRFQEGKLESGNLIYDFQGVVSLGEFSSVPDEVREQITDAIS
TIQTQGFPHHEER"
complement(9014..9187)
/gene="BH2890"
complement(9014..9187)
/gene="BH2890"
/note="BH2890"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06609.1"
/db_xref="GI:10175512"
/translation="MNVKRAIFGIIFAIITISTYILFHNVLVFSDFSVIALVCGF
LVERLFMKWRHAK"
complement(9233..10138)
/gene="BH2891"
complement(9233..10138)
/gene="BH2891"
/codon_start=1
/transl_table=11
/product="UDP-glucose 4-epimerase"
/protein_id="BAB06610.1"
/db_xref="GI:10175513"
/translation="MNRVLITGSGFIGVLCLEALLGQGDIVTIDHCNEMRKEEYE
EQRLMGRNALPHFIECSLEHPELSVTQOVDTYHLAHPKREQLWTRLPENQENTL
RNRIDTAQAKEGTRILFVSSVEYGERSGTIEQSPTRPTSPYGIIVKLAERKLQKE
AKRGLDYMILRPTIYFPWQDMTYOQLLGRKEVKDDQSTDALYISDAINGLLL
AKGKHTNETIOLSSGNVEMQKRGAKHLLKNKSNWSPSTPLSPILSNEKAKQLIGFKVI
TPLNGLKQREHVTOQWRKWSFLH"
complement(10170..10916)
/gene="BH2892"
complement(10170..10916)
/gene="BH2892"
/note="BH2892"
unknown conserved protein in bacilli"
/codon_start=1
/transl_table=11
/protein_id="BAB06611.1"
/db_xref="GI:10175514"
/translation="MIDLRRTFQLEDQWCVIVLPARPNGFACMLFGDITHYVEQETCS
WQHPPREQFIRGLLHHGYTITSNFFGRHWGSEQATKLALRYHLISIKREILNPSIH
LIGEGMALLANVMEEMPDRIKSGFFINPCLDLYSYDQEKANKLFYKRELIEMAAA
HEVSTEDVEFFIYQVFNRRYTNLETPMHIFHEMNERRYPVQSHSKPFLKAEKRNRPV
TITLRNAQSFSSLDIPAYQFFRKHEKLS"
complement(11473..11652)
/gene="BH2893"
complement(11473..11652)
/gene="BH2893"
/note="BH2893"
unknown conserved protein in bacilli"
/codon_start=1
/transl_table=11
/protein_id="BAB06612.1"
/db_xref="GI:10175515"
/translation="MGQNRQFIPGQKAPNNGVVEIGETGDMVTDPREVKLEVGDRFP
ETTNKDRKWTYKRKP"
complement(11705..12664)
/gene="argf"
complement(11705..12664)
/gene="argf"
/note="BH2894"
/codon_start=1
/transl_table=11

/product="ornithine carbamoyltransferase"
/protein_id="BAB06613.1"
/db_xref="GI:10175516"
/translation="MNHSEKVISLGRDLLTLDDYTPPEVQQLLTQALLELKOKAKNG
EPTPYLTGSLGMIFENASTRTVRVSFEGMTQUGHALFLSPKDLQIGRGEPIKDTAN
VLSRYVDAMIRTNSHESVEELAHYATVPVYNALTDAYHPCQALADALTLEKKEITLI
GKTLAYIGDGNVSHSLAIGATGMDTVATPKGYEVDQEIFQRATEAAKGTGATLV
QNTPOVAENADAIYTDVWASMGYEAQESEREVEVPQYQVNDQLTLAKKDYVSFLHC
LPAHRCGEVTASVIDGPHSAIYDQAEENRHLHAQAKVLTALLIGE"
complement(12689..15832)
/gene="carB"
complement(12689..15832)
/EC_number="6.3.5.5"
/note="BH2895"
/codon_start=1
/transl_table=11
/product="arginine specific carbamoyl-phosphate synthase
large chain"
/protein_id="BAB06614.1"
/db_xref="GI:10175517"
/translation="MPKRTDIOQSVLIGSGPIVIGQAAEEDYAGAACLALREEGIQV
ILVNNPAIVMTDEACADVIFPLTVASVKNIIERERPDGLLATLGQGTGLNLMKL
EAGILEIYNVELLGTPMESIKKGEDREAFRQLMHELHEVPSEIVHSVAEADVAFAN
TVGYPIIVPAYTLGGAGGIAESEALIRIVKGGLELSP IQOCLIEKSIAGKEIEY
EVMRDSNDCITVCNMENDPVGHHTGDSITVWAPSOTLTDOEYMLRSASLKIIRTLIG
IVGGCNIQFALDDPSKOYVLIENVPRVSRSSALASKATGYIARMAAKLSLGLHLEL
KNPVEDTYASEPESLDYVVVFRPMPDKLVHVRRELGTQMKATGEVMAIERNLKAG
LQKVRSLKITHGLSLPSQWEDSELWIVVKADDRFFALLELRGRTVLEAIEHAD
QTKIDRFLTSAKMLTLEKETAQSDDDITSELTYSKKYGFSDMLASSWGVGLAD
VRHTRKALGVPSYKMGVCAAEFAKTPYYSSWTGENDLLLPKAKERVLIIIGSGP
IRIGQIEFDYDYSVGHAKSLARNAFEAIINNPNETVSITDYETADRLYFPLAVEVDL
NIVEVENDHVIVQLGSGTAIGLTGLEAGVSILGTOTDVIQLEDRERREYFEMRSV
EVPHIEKTAETKEELLKAOSIGYPIILRPYSVIGOGMFIASNOEELAAFCEDKNH
SVTFPIIDVAYPGVEFVDVLDGSDITIPGMEFHEVKAGVHSGDSMAVTPPTTLEA
KNQQAINTRQIAKGMAYKGLFNIQVLYDEELYVIEYNAGRSRIVPIFSKATSLPL
ITYITDVLFGKTIAGELGASGYKESPYTVKAPVSYQKLAGLDPLLEAKEMKSTGEL
IATSLPAPSAKRAKFAWGEQTPALFRKKGSVFCQVDRAVDTQWQPLLRQLKKEKYSV
VTEAMSFSLEWASADAICLVSPAPQKTKQNRBEALKQRTVTVVSDLTATEFKMEIC
LEVKGDEPFLPDVNM"
complement(15825..16904)
/gene="carA"
complement(15825..16904)
/EC_number="6.3.5.5"
/note="BH2896"
/codon_start=1
/transl_table=11
/product="arginine specific carbamoyl-phosphate synthase
small chain"
/protein_id="BAB06615.1"
/db_xref="GI:10175518"
/translation="MKAILVLTGETFAGEIAINAKEDVYGEVVFTGMTGYQVLSDP
SKFGQLVVTFYPLIGNINADSFESQAGLIVSEQEGHHYEAOTOSLQOQCDQ
HQPLLTGIDTRAVVRIREQGDMPAVITDLSRVSEFENVPLSERALNDEVSTKKVE
TFEGNGPHVLIIDYGVKQISLNSLLKRDQCVTVPYDVSFVQVSELNPCDLFLSNGPG
DPKQIEDRLPTIYRLASTYPSLIGICIGHOLALAFGADTEKLFREGHGANQPVNLKS
NRVWTSQNHSYVVKEDSLQRTDWDATFRKNINDSGSIEGLTHKYLPIQTQVQFHPHFG
PSDSEIFHSFIDDDVAAGKRETYA"
complement(16980..18134)
/gene="argD"
complement(16980..18134)
/EC_number="2.6.1.11"
/note="BH2897"
/codon_start=1
/transl_table=11
/product="N-acetylornithine aminotransferase"
/protein_id="BAB06616.1"
/db_xref="GI:10175519"
/translation="MSHLFPTYAKWDVTTISGKGTKLYDNQGNNEYLDYFVSGIATVCNLG
HCHPKVVAAEQHLFSHVNLFHIPQENVAALLTEQSDMADVFCNCSGAENDEAA
IKLARKATCKHEIVTFTQSFHGRTLCTMSATGQDKIKTGTGPMLETFHHVPFNDIAAL

KOVNEQTAAIVLEIVIOEGGVNLIIDPEFAASVNVHVCQEHGILLIIDEIQTGIGRTGT
 AFAPQOYELTDDIITVAKGLNGFPVGMALGKQHLIDAFSAGSGSTFGGNPLAMAAA
 QAVLTFEPQFNQAOVQEGKOLLGSLNEALSGLEIVKEIRGNGLIYGLIEQESGAPP
 IKOLREKGLLVNAGPNVIRLLPPLVYVTSSELHEAVTQLREVLDQASVHA"
 complement (18147. .18929)
 /gene="arg8"
 /complement (18147. .18929)
 /gene="arg8"
 /EC_number="2.7.2.8"
 /note="BH2898"
 /codon_start=1
 /transl_table=11
 /product="N-acetylglutamate 5-phosphotransferase"
 /protein_id="BAB06617.1"
 /db_xref="GI:10175520"
 /translation="MGDIVIKCGSVLDSLSDFTSVRRLQAEAGAKPIIVHGGGPA
 INGLAKLEVETTFVDFGLRKTTSVLSVAEMVLCGGQTKKVVRRKLQKAGVKISGLSGS
 DGLVVRVKAIDEALGLYGVPEPVHNGELLHMLVADGEVPLAPIGVNDPEFEPYNVAD
 TVAGAVATLQAKELIFVTDVAGILKDDLVLPALTPPEVESLIEEGT IYGGMIPKVR
 AHSLTGLEAVSIVNGNDVFFAESGKLICTTIKKQVHSQG"
 complement (18958. .20193)
 /gene="argj"
 /complement (18958. .20193)
 /gene="argj"
 /note="BH2899"
 /codon_start=1
 /transl_table=11
 /product="glutamate N-acetyltransferase/amino-acid
 acetyltransferase"
 /protein_id="BAB06618.1"
 /db_xref="GI:10175521"
 /translation="MNVINETANVLKLTGVSITSAGFSAVGHTGVKRRKDLGAIV
 CEVPASSAAVYTLNKVQAAPLKVTOESIAVEGLQAMINSGSTANACTGKRGDLDAIT
 MRVAGAEPIPEHYAVYVSTGVIGFELPMDVITINGRLKPEATIEGAHAFNEAIIIT
 TDVKEHTQYQTVNGKTVTVGGVAGKSGMIHPNMATMLSFVTTDANIHDHGLQGALS
 AITNETFRITVDGDTSTNDVMVVMASGLAENEALTPEHPDWMANFYKALQACEDLAK
 QTARDEGATKIEVEVTGAANDQEAQVAKQIVGSDLVKTAIYGADANNWGRITCAIG
 YSGCEVNETIDIAIGPIVITLKQSEPTGFSSEATAYLKEADPVKISVNLHIGNGTGK
 AWGCDLTYDVVRINAGYR"
 complement (20208. .21245)
 /gene="argc"
 /complement (20208. .21245)
 /gene="argc"
 /EC_number="1.2.1.38"
 /note="BH2900"
 /codon_start=1
 /transl_table=11
 /product="N-acetylglutamate gamma-semialdehyde
 dehydrogenase"
 /protein_id="BAB06619.1"
 /db_xref="GI:10175522"
 /translation="MKVGIVGATGYGGLLEIRLLSQHPNWEELICYSSSQEGLPLDDM
 YPHLSGNVPWTLKAIDPEAIRADVDTVFLATPPGVSGELTDPDLVGLKVIDLSGDLR
 INEPDVEAWYKRAAPVGTIRGAVYGLTEWQORDEIAAAQIIANPCVYPTAVLLGLAP
 LVQQWIDPKRIIDAKSGTSGAGNPSQITHFSEMMENFKIYQVQNKHHTPEIQOL
 RNWDESQVNGVSTHLLVPMVRGIMATIVAEANOALSEKELREHFAVYVETSPFVHVS
 EGRYPATKEVFGSNRCHIGVTVDTERTNRTTVSVVIDNLVKGAGAGAIQNYNVNMYEE
 TGLGEGSLPY"
 complement (21282. .21596)
 /gene="BH2901"
 /complement (21282. .21596)
 /gene="BH2901"
 /note="BH2901"
 unknown
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06620.1"
 /db_xref="GI:10175523"
 /translation="MVSALFTFSKEMALLKINGETLYKQAIKVLKSDHHSFHSV
 AMDLSGLRYRVNNEHRRERELNMKTSALIVCFHRELNKTKLIYKNTVEFIFIQFYNQK
 QR"
 complement (21668. .21997)
 /gene="BH2902"

CDS

complement (21668. .21997)
 /gene="BH2902"
 unknown
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06621.1"
 /db_xref="GI:10175524"
 /translation="MGSGLRLYREPDAFQDMLGVLYWQGEVSLVDAEITRNGELVH
 KAEFYEEAILTEHQSEDFIHSFQSMFTNKNDYIQLLIHWGDSGTHEDIMVLSPKNH
 YFVMPRY"
 complement (22273. .23958)
 /gene="BH2903"
 complement (22273. .23958)
 /gene="BH2903"
 /codon_start=1
 /transl_table=11
 /product="oligo-1,6-glucosidase"
 /protein_id="BAB06622.1"
 /db_xref="GI:10175525"
 /translation="MSTQWKKESVYQIYPRSFQYNGDGIQIPGISRLDYLKTIG
 VDVIWLSPYVDSNDNGYDIRDYKAIMDEFGTMADWETLLAEITHRGMKLIMDLVNV
 HSDDEHAFVESRKSXDNPYRDEV IWRPGDKGKEPNNAWNSFGSANTYDETTEYYIL
 HLFSSKQPDLMNENPKLREKIYEMMTWLDKGDGFRMDVINFSKVDGLPDAEPQPG
 QPYVSSGNYFMNPNHITIQEHENHVLQHYDLMTVGMFPGVTLELAQLYTGEERNE
 NMVYQFEHVGLDQGPNGKDLKPLEKDLKASLRQKQDIGNWSLYNNHDPRI
 VSRFGDDQSYVESAKMLATLLHCKMGTPIYQCEEIGMTNVRFDSIQYQDITLNM
 YKERAQGVPHETLMASIHAKGRDNATPMQWDETGHGFTDGTGTPWLEVPNPKYKEINV
 KOALKPNSTIFYHYOKIQLRKHEALIVHGSYDLILLEDDEPEIYAKKTYNQTLVVC
 NFYGRITDECPAEVVLSEPTLLLSNYDEENGSYTSFRURPYEARVYLKGRN"
 complement (24182. .24787)
 /gene="BH2904"
 complement (24182. .24787)
 /gene="BH2904"
 /note="BH2904"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06623.1"
 /db_xref="GI:10175526"
 /translation="MKRWMTIAVLALMLGCGNWLDMGNSNEATDPAFDISEAGLTVP
 DPEYVNOEGDPFCTEOLKGNVWLVNMFPHCGVCVCHIMFPNWNVLQDAMIIEGLDKVY
 VSFTVDPRDTPPELLAKYGVENGVNAGDFEYFV7GYSYSGEITELAKATAFASIVQEDPEN
 EDLITDFFLVDDQGLVIRRYDGMATDQAEIINDLKATVE"
 complement (24804. .25718)
 /gene="BH2905"
 complement (24804. .25718)
 /gene="BH2905"
 /note="BH2905"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06624.1"
 /db_xref="GI:10175527"
 /translation="MDQLESNCFRALWTPELIILIANVGFVFMATKWRERFTNSE
 PVPKRIKTVLGLITALYIGWGSFPYVAGHLMITFMAQMFAYFATVPLFLIGTPAW
 VESALFEKLRPWSEQFAKIANNPVALLELNLGSLYHVPFMSQSPILHSLYQ
 FLLWYGAELMMWHMIAPITSTKSLDLRIRIGYIFANGILITPACALIIIPAPQLYQY
 TNPAIVAMVAYCLPPGSSVPLSVENGPDSTFLETMRDQQLAGVLMKIMQETTYGFT
 IGVYFKQWLSKEKQDQGALTISDVPYK"
 25877. .26686
 /gene="BH2906"
 25877. .26686
 /gene="BH2906"
 /note="BH2906"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06625.1"
 /db_xref="GI:10175528"
 /translation="MPORHIALDLDDGTLTDNKTISMKTOTICKAREAGHIVVIST
 GRPYRASIQYQELQDIAIVNFGNAGVHPKDSFGTYHHPLELSTARQVIECEAF
 DVSNIWVEIDDIYLYRDELFIQFTTEGQGVPERGNLLKLKRDPTCVLIHPKDDHV

gene

CDS

complement (22273. .23958)
 /gene="BH2903"
 complement (22273. .23958)
 /gene="BH2903"
 /codon_start=1
 /transl_table=11
 /product="oligo-1,6-glucosidase"
 /protein_id="BAB06622.1"
 /db_xref="GI:10175525"
 /translation="MSTQWKKESVYQIYPRSFQYNGDGIQIPGISRLDYLKTIG
 VDVIWLSPYVDSNDNGYDIRDYKAIMDEFGTMADWETLLAEITHRGMKLIMDLVNV
 HSDDEHAFVESRKSXDNPYRDEV IWRPGDKGKEPNNAWNSFGSANTYDETTEYYIL
 HLFSSKQPDLMNENPKLREKIYEMMTWLDKGDGFRMDVINFSKVDGLPDAEPQPG
 QPYVSSGNYFMNPNHITIQEHENHVLQHYDLMTVGMFPGVTLELAQLYTGEERNE
 NMVYQFEHVGLDQGPNGKDLKPLEKDLKASLRQKQDIGNWSLYNNHDPRI
 VSRFGDDQSYVESAKMLATLLHCKMGTPIYQCEEIGMTNVRFDSIQYQDITLNM
 YKERAQGVPHETLMASIHAKGRDNATPMQWDETGHGFTDGTGTPWLEVPNPKYKEINV
 KOALKPNSTIFYHYOKIQLRKHEALIVHGSYDLILLEDDEPEIYAKKTYNQTLVVC
 NFYGRITDECPAEVVLSEPTLLLSNYDEENGSYTSFRURPYEARVYLKGRN"
 complement (24182. .24787)
 /gene="BH2904"
 complement (24182. .24787)
 /gene="BH2904"
 /note="BH2904"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06623.1"
 /db_xref="GI:10175526"
 /translation="MKRWMTIAVLALMLGCGNWLDMGNSNEATDPAFDISEAGLTVP
 DPEYVNOEGDPFCTEOLKGNVWLVNMFPHCGVCVCHIMFPNWNVLQDAMIIEGLDKVY
 VSFTVDPRDTPPELLAKYGVENGVNAGDFEYFV7GYSYSGEITELAKATAFASIVQEDPEN
 EDLITDFFLVDDQGLVIRRYDGMATDQAEIINDLKATVE"
 complement (24804. .25718)
 /gene="BH2905"
 complement (24804. .25718)
 /gene="BH2905"
 /note="BH2905"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06624.1"
 /db_xref="GI:10175527"
 /translation="MDQLESNCFRALWTPELIILIANVGFVFMATKWRERFTNSE
 PVPKRIKTVLGLITALYIGWGSFPYVAGHLMITFMAQMFAYFATVPLFLIGTPAW
 VESALFEKLRPWSEQFAKIANNPVALLELNLGSLYHVPFMSQSPILHSLYQ
 FLLWYGAELMMWHMIAPITSTKSLDLRIRIGYIFANGILITPACALIIIPAPQLYQY
 TNPAIVAMVAYCLPPGSSVPLSVENGPDSTFLETMRDQQLAGVLMKIMQETTYGFT
 IGVYFKQWLSKEKQDQGALTISDVPYK"
 25877. .26686
 /gene="BH2906"
 25877. .26686
 /gene="BH2906"
 /note="BH2906"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06625.1"
 /db_xref="GI:10175528"
 /translation="MPORHIALDLDDGTLTDNKTISMKTOTICKAREAGHIVVIST
 GRPYRASIQYQELQDIAIVNFGNAGVHPKDSFGTYHHPLELSTARQVIECEAF
 DVSNIWVEIDDIYLYRDELFIQFTTEGQGVPERGNLLKLKRDPTCVLIHPKDDHV

gene

CDS

complement (24182. .24787)
 /gene="BH2904"
 complement (24182. .24787)
 /gene="BH2904"
 /note="BH2904"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06623.1"
 /db_xref="GI:10175526"
 /translation="MKRWMTIAVLALMLGCGNWLDMGNSNEATDPAFDISEAGLTVP
 DPEYVNOEGDPFCTEOLKGNVWLVNMFPHCGVCVCHIMFPNWNVLQDAMIIEGLDKVY
 VSFTVDPRDTPPELLAKYGVENGVNAGDFEYFV7GYSYSGEITELAKATAFASIVQEDPEN
 EDLITDFFLVDDQGLVIRRYDGMATDQAEIINDLKATVE"
 complement (24804. .25718)
 /gene="BH2905"
 complement (24804. .25718)
 /gene="BH2905"
 /note="BH2905"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06624.1"
 /db_xref="GI:10175527"
 /translation="MDQLESNCFRALWTPELIILIANVGFVFMATKWRERFTNSE
 PVPKRIKTVLGLITALYIGWGSFPYVAGHLMITFMAQMFAYFATVPLFLIGTPAW
 VESALFEKLRPWSEQFAKIANNPVALLELNLGSLYHVPFMSQSPILHSLYQ
 FLLWYGAELMMWHMIAPITSTKSLDLRIRIGYIFANGILITPACALIIIPAPQLYQY
 TNPAIVAMVAYCLPPGSSVPLSVENGPDSTFLETMRDQQLAGVLMKIMQETTYGFT
 IGVYFKQWLSKEKQDQGALTISDVPYK"
 25877. .26686
 /gene="BH2906"
 25877. .26686
 /gene="BH2906"
 /note="BH2906"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06625.1"
 /db_xref="GI:10175528"
 /translation="MPORHIALDLDDGTLTDNKTISMKTOTICKAREAGHIVVIST
 GRPYRASIQYQELQDIAIVNFGNAGVHPKDSFGTYHHPLELSTARQVIECEAF
 DVSNIWVEIDDIYLYRDELFIQFTTEGQGVPERGNLLKLKRDPTCVLIHPKDDHV

gene

CDS

complement (24804. .25718)
 /gene="BH2905"
 complement (24804. .25718)
 /gene="BH2905"
 /note="BH2905"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06624.1"
 /db_xref="GI:10175527"
 /translation="MDQLESNCFRALWTPELIILIANVGFVFMATKWRERFTNSE
 PVPKRIKTVLGLITALYIGWGSFPYVAGHLMITFMAQMFAYFATVPLFLIGTPAW
 VESALFEKLRPWSEQFAKIANNPVALLELNLGSLYHVPFMSQSPILHSLYQ
 FLLWYGAELMMWHMIAPITSTKSLDLRIRIGYIFANGILITPACALIIIPAPQLYQY
 TNPAIVAMVAYCLPPGSSVPLSVENGPDSTFLETMRDQQLAGVLMKIMQETTYGFT
 IGVYFKQWLSKEKQDQGALTISDVPYK"
 25877. .26686
 /gene="BH2906"
 25877. .26686
 /gene="BH2906"
 /note="BH2906"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06625.1"
 /db_xref="GI:10175528"
 /translation="MPORHIALDLDDGTLTDNKTISMKTOTICKAREAGHIVVIST
 GRPYRASIQYQELQDIAIVNFGNAGVHPKDSFGTYHHPLELSTARQVIECEAF
 DVSNIWVEIDDIYLYRDELFIQFTTEGQGVPERGNLLKLKRDPTCVLIHPKDDHV

gene

CDS

complement (24804. .25718)
 /gene="BH2905"
 complement (24804. .25718)
 /gene="BH2905"
 /note="BH2905"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06624.1"
 /db_xref="GI:10175527"
 /translation="MDQLESNCFRALWTPELIILIANVGFVFMATKWRERFTNSE
 PVPKRIKTVLGLITALYIGWGSFPYVAGHLMITFMAQMFAYFATVPLFLIGTPAW
 VESALFEKLRPWSEQFAKIANNPVALLELNLGSLYHVPFMSQSPILHSLYQ
 FLLWYGAELMMWHMIAPITSTKSLDLRIRIGYIFANGILITPACALIIIPAPQLYQY
 TNPAIVAMVAYCLPPGSSVPLSVENGPDSTFLETMRDQQLAGVLMKIMQETTYGFT
 IGVYFKQWLSKEKQDQGALTISDVPYK"
 25877. .26686
 /gene="BH2906"
 25877. .26686
 /gene="BH2906"
 /note="BH2906"
 unknown conserved protein
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06625.1"
 /db_xref="GI:10175528"
 /translation="MPORHIALDLDDGTLTDNKTISMKTOTICKAREAGHIVVIST
 GRPYRASIQYQELQDIAIVNFGNAGVHPKDSFGTYHHPLELSTARQVIECEAF
 DVSNIWVEIDDIYLYRDELFIQFTTEGQGVPERGNLLKLKRDPTCVLIHPKDDHV

gene

CDS

complement (21668. .21997)
 /gene="BH2902"

/gene="BH2912"
 /note="BH2912
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06631.1"
 /db_xref="GI:10175534"
 /translation="MAETTPSLSIIEYVQHLNKKFQSYAERGKRYHEGKVINFSVQ
 SNRIEAOVDGTHLYHVTVLDDFGLSNTCTPVEDYCKHMAAALYEQWQHOVLANIWD
 KNIPOVIDTLHPVLSPSTYROLISRAYLVAQVEELMEAMKTAVKQHFSDAEAAVYI
 YIIQAMGNATSYQIYERQERRRFSFQGLQEVIIQALPALSDDLLRLAKAMRFSFL
 SHENTYIGWESLMREFLPVPSRLYQRLSDLLDATYEQPLGKGRSLVKSYPFLAEG
 HGLLALQQLLEGHTIRPDLVPHMTWMTNHWTTMKAFHMLFPRKESAPFLVLPQPYD
 EMEVFATYGLSSQSNWALQQPSYQRLSLTKNWSKRSALQELLPALKNQLE
 QMPSRTAYTKLLTLEKRAEAGANYLQYETEPFLRPEVNNELLOMQIQRNRELLFPVY
 HQFVLRLEIEKKSRRPHYEEAVHFLILHCLYKETEDKEREFHQYIRRLKYTKSYVALLE
 ELRQLDS"
 31788..34775
 /gene="BH2913"
 31788..34775
 /gene="BH2913"
 /codon_start=1
 /transl_table=11
 /product="SNP2 helicase"
 /protein_id="BAB06632.1"
 /db_xref="GI:10175535"
 /translation="MIPNIIVHGGWFKDHFITWGEQASSSSGRIVRFEVPFLYSPFE
 LRLAFRHPHSPFGSFLQGOAVLEPLAERAFHSPAGTTTIYQAEPSMKSRFPQVE
 GIMLSADELFDYGSFLSKGSSSDWVLADDTFHFHLHFFLYIRQIRNGARSPSAQ
 TWVNNPNKAAEAPEVCLALTENAPLIOTTEEPVEHQLOAIGELHJTDKWRLL
 QTDKGRDLFSYQSLHPSVKRTLEQMPYGEGETEETHSEKVDROHFLAAGVKRPOA
 PQCAFVLPKITGTDWTLSCVVDKNPSIIVEMEELSQGHPPWRMNPVPLKKRL
 PAVTDRPLQSLIKSSSPVIKLSNEEAYHLFTTYDEYLYKELGFLVPRWLTORRPLK
 VKLKMDAHLHSEVEPLLNNQSVASQYELAIGDNVSLSEFQTVFVDRNPFLQNGE
 WIADPSLAKLRAYLNLKLEHGQTTYLDARHLEQSEGSFQEAWELENSAIEQKI
 KOLFOONVLRPAPASQATLRPYQEEGASLHLHLETFGGCLADDMGLGKTQVTIS
 YILVLEHGQNNKPPFLFPCPTSLITNNWHECKTFAPSLNYYVHHQQHQENETAMR
 EADIYSSYLAYKOLDQWKDIENGLIIDEAQIKNVDTKQAVKSIKRAAHIALT
 GPTENRLKELSIMDVNLPSYKLSLTHTRTFIKPIEREKDEARQEQLQLOIPLLL
 RKKSDEALALQLOPEKREHVRVSLSEQAALYQAVVDNNHQLGDTIMERRALLLK
 TITKLKQINPAHFLKDNDAHQSEKWEILLTLEQIMDRQEKMLITIQFKEMGHL
 MQDAFQSGHTPIPIPLHGSLSRQROEAVERFONDRLEPIFILSLKAGGVGLMTAAN
 HVIYIDRWNPVAVENQATRAYIRIGQTADVTYVHKLMTGTELEERTIHOMLSKQALAEQ
 ILTAGENQLTELTDDELLALRL"
 34789..34971
 /gene="BH2914"
 34789..34971
 /gene="BH2914"
 /note="BH2914
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06633.1"
 /db_xref="GI:10175536"
 /translation="MAADDRFSHYWRLLKQLPKIDTPIYEDESLLKEEWLKLKRAIQ
 EAKSAFPNTEKSED"
 complement(34984..35490)
 /gene="BH2915"
 complement(34984..35490)
 /gene="BH2915"
 /codon_start=1
 /transl_table=11
 /product="intracellular proteinase inhibitor"
 /protein_id="BAB06634.1"
 /db_xref="GI:10175537"
 /translation="MRRCITLLIALGLATLSACQGTSSPSGDTNGDAEELKEEES
 EYNGVNVSAETVGDQIEATITLINTDQDVHFQFSSGOYELILKNAGDTVYVYS
 DOKMFTMALIDETLAGESKWTWESIPAGDLESGTYTLLEAVVVAADVGSVPKPSY
 KAQTTVIK"
 complement(35526..36407)
 /gene="BH2916"
 complement(35526..36407)
 /gene="BH2916"

gene
 CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS

/note="BH2916
 unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06635.1"
 /db_xref="GI:10175538"
 /translation="MDGKTVNANDYICEDMMKOTMMNELKKIIVVVGIGAVFLAASLNF
 FLIPADVFASGFTGVAQILATILPLSTGILLLLNIPVAILGWLKLGKAEFTVYSFJSV
 ALTLFLEVPVQAFSPDILLNAVFGVAAIGAGITLVKVGASGGDIDIIALLISRV5
 DRVGIYFFLLNGLIYVASGFLDAEKALYTLVTLYVTSRIIDAIRHRVKLTAMIVT
 KPTELRKAHERITRITRMPAKGGYLEDDEKVELMIVLTRYELYDLRHRVQ5VDPEA
 FTVIHTEGVYGLFRKT"
 complement(36359..37138)
 /gene="BH2917"
 /note="BH2917"
 complement(36359..37138)
 /protein_id="BAB06641.1"
 /db_xref="GI:10175544"
 /translation="MNFPMFGLKDTFTKVEEVEIVRLYVEMERKVRHRCPKQKR
 TKVHDYRIQIKHLKWFERQTIQFYARRRYVCAGKRFAENNSFVRYQRTSIEWQ
 AVAIRAIGKTFKETAENYGSATTIVRREDQISSSEIRPEVELPAIIAIDEYKGDTK
 EGKYQLIIDAVTKPELDILPNRYKRTIKDYLRRHGSKVQVIMDMSPSKAAVQOEAL
 GRPVIITADRFHFCRYIYWALDGVRRVKQDHFAYDRKKKRMKHVFKENSLTEEPER
 WYLDRYLKSTELQKAYELKEAYQDMFKSKSGATQIERVKKELKAIYKLVKNSGIP
 EMCRAICTENMQTEILNSPAYGYSNGFLEGINNSTKVLKRNAGFRFSYERFRAKILL
 THYKRIQVHIG"
 complement(41076..42110)
 /gene="BH2923"
 /note="BH2923"
 complement(41076..42110)
 /protein_id="BAB06642.1"
 /db_xref="GI:10175545"
 /translation="MAVTIKDVAKLAKVAPSTVSRYVIANPNRISQRTKERVAMEOL
 GYPNPNASLVNQSKTIGIIIPSSARTAQPNFFPEVIRGITTKAYEKGILYLST
 GQTEELLAEVKMVSRRVDGILLIYAKODDKMPYLIEENFFVLVGRPYDDQKLE
 LV5FVNNDVKAAMVTEYILLGHRKIGFTIGNLEYITVDIHKEGYRRALANAGIPF
 REYVYVYDEVEGQGEAVIELMSKEPTALIVADDLMALGVRLMDMDVSVPEEM
 SIISFNAMISELSPPLTVDIHFIDGLGQATECLIERKMKDPERPPAQIIPHKLK
 ROTCRLLKS"
 complement(42224..43066)
 /gene="BH2924"
 /note="BH2924"
 complement(42224..43066)
 /protein_id="BAB06643.1"
 /db_xref="GI:10175546"
 /translation="MNKVKSRLEVTAIYFLVLMGIVILYPLLWTVGLSNPGTSLF
 SSRMIPETISFRHYEWLFFDPRSNYLQWKNTLIVASVTSVCSTFLVALTAYAF5RYR
 FVGRYGLXGFLLLQMFVPMARVALYIILNTVNLDTLLGLILLIYVGTSPKNNAFLY
 KGFDTIPRELDESAKLDGHRIFFTIIMLPIAKPLI5AVLNFPMSPMDFILPRI
 ILRSPENYTLALGLFNVDQFANNFRFAAGAILIAPIATVFLFLQRYLYLSGLTTG
 ATKG"
 complement(43066..44358)
 /gene="BH2925"
 /note="BH2925"
 complement(43066..44358)
 /protein_id="BAB06644.1"
 /db_xref="GI:10175547"
 /translation="MEKEVREKAATNPKLALLSTIIPGLQOLYNKRYLKGSVFFILF
 AAFIIVHDFLNGIWGIMTGLTLEVRDSDRMLLAGQIMALITVPAVGLYLNLNDRDA
 YDKAKRIQGWRIPIPLRESFOQAWDKPMLIGPLIGLIFITVVPFLMGLAFTN
 YNLNAPPNLLDVGFDNFVOLVTPIWRETFESVSWTLIWTFFVATSCQIALGLFL
 AMVNDPRILKFKKFIETILLPWA5PAFTIILFSAMFNDFGANRDIIDPLFGHGA
 PMLTDFWSRFALIMIQTWLGFPPFALFTGLVLSISSD5WYEAADMGANRLQKFRHI

/translation="MPOGKWFRFGYGVTLVLLIYVLSLVDWIFRPVIVVQTLFAPV
 IIRAGIYLFPRPIVGLSKKMNRLGSLILILFLMIGIAGTAFSLJIVGPELQRFQSLYN
 NLPTLINQVRDFTIALOSNELFNFOESETFOPEEMISRFVNSLNDIIGTIGANITGR
 IGALANFFIILVLVPFTIYLLKKEGKLPVQLLKFNKKQKHGRLILADMDKMLSSY
 IQGQIVSFCVGLICYIAYLIIIGLDYALIIIALVAMFNVPFVPGPWIGTAPVIVAFI
 D5PLTALIVIVVVVIOQIESNLISPOVMKKLAIHPITITIIILLVAGRFAGVVGMLL
 AVPTYAGKRVIVTHSYRLKLRKQPV"
 complement(39701..40915)
 /gene="BH2922"
 /note="BH2922"
 complement(39701..40915)
 /protein_id="BAB06641.1"
 /db_xref="GI:10175544"
 /translation="MNFPMFGLKDTFTKVEEVEIVRLYVEMERKVRHRCPKQKR
 TKVHDYRIQIKHLKWFERQTIQFYARRRYVCAGKRFAENNSFVRYQRTSIEWQ
 AVAIRAIGKTFKETAENYGSATTIVRREDQISSSEIRPEVELPAIIAIDEYKGDTK
 EGKYQLIIDAVTKPELDILPNRYKRTIKDYLRRHGSKVQVIMDMSPSKAAVQOEAL
 GRPVIITADRFHFCRYIYWALDGVRRVKQDHFAYDRKKKRMKHVFKENSLTEEPER
 WYLDRYLKSTELQKAYELKEAYQDMFKSKSGATQIERVKKELKAIYKLVKNSGIP
 EMCRAICTENMQTEILNSPAYGYSNGFLEGINNSTKVLKRNAGFRFSYERFRAKILL
 THYKRIQVHIG"
 complement(41076..42110)
 /gene="BH2923"
 /note="BH2923"
 complement(41076..42110)
 /protein_id="BAB06642.1"
 /db_xref="GI:10175545"
 /translation="MAVTIKDVAKLAKVAPSTVSRYVIANPNRISQRTKERVAMEOL
 GYPNPNASLVNQSKTIGIIIPSSARTAQPNFFPEVIRGITTKAYEKGILYLST
 GQTEELLAEVKMVSRRVDGILLIYAKODDKMPYLIEENFFVLVGRPYDDQKLE
 LV5FVNNDVKAAMVTEYILLGHRKIGFTIGNLEYITVDIHKEGYRRALANAGIPF
 REYVYVYDEVEGQGEAVIELMSKEPTALIVADDLMALGVRLMDMDVSVPEEM
 SIISFNAMISELSPPLTVDIHFIDGLGQATECLIERKMKDPERPPAQIIPHKLK
 ROTCRLLKS"
 complement(42224..43066)
 /gene="BH2924"
 /note="BH2924"
 complement(42224..43066)
 /protein_id="BAB06643.1"
 /db_xref="GI:10175546"
 /translation="MNKVKSRLEVTAIYFLVLMGIVILYPLLWTVGLSNPGTSLF
 SSRMIPETISFRHYEWLFFDPRSNYLQWKNTLIVASVTSVCSTFLVALTAYAF5RYR
 FVGRYGLXGFLLLQMFVPMARVALYIILNTVNLDTLLGLILLIYVGTSPKNNAFLY
 KGFDTIPRELDESAKLDGHRIFFTIIMLPIAKPLI5AVLNFPMSPMDFILPRI
 ILRSPENYTLALGLFNVDQFANNFRFAAGAILIAPIATVFLFLQRYLYLSGLTTG
 ATKG"
 complement(43066..44358)
 /gene="BH2925"
 /note="BH2925"
 complement(43066..44358)
 /protein_id="BAB06644.1"
 /db_xref="GI:10175547"
 /translation="MEKEVREKAATNPKLALLSTIIPGLQOLYNKRYLKGSVFFILF
 AAFIIVHDFLNGIWGIMTGLTLEVRDSDRMLLAGQIMALITVPAVGLYLNLNDRDA
 YDKAKRIQGWRIPIPLRESFOQAWDKPMLIGPLIGLIFITVVPFLMGLAFTN
 YNLNAPPNLLDVGFDNFVOLVTPIWRETFESVSWTLIWTFFVATSCQIALGLFL
 AMVNDPRILKFKKFIETILLPWA5PAFTIILFSAMFNDFGANRDIIDPLFGHGA
 PMLTDFWSRFALIMIQTWLGFPPFALFTGLVLSISSD5WYEAADMGANRLQKFRHI

/codon_start=1
/transl_table=11
/protein_id="BAB06654.1"
/db_xref="GI:10175557"
/translation="MKYCIHNTRLIIIGIADRVIEINGSILFDQDKILKVSASHSITDADV
FIDGTTFSVPLGLIDVHVHLGMDCSVDPFDMTEADFTSFIAIKOGLQFLRAGITV
VRNGTKNNVDIIYKKAVESGLVIGPRVYAAGQPLIMTGGHQPMGTEVDCQKDEVIKA
TRQRIKEGADILIKIMATGQVLTCKGNHPGAVHFNDELKCIQCEAAKASLTTAAHAIGL
NGVNAIKAGWNTIEHGYLYDKETVDPMRDGLIYVLTPLAPTLLOKNSHIIPOHMD
KIEIEKHKRSFRLALRGVKIIITAGDAGTGFNNPGOLVVEELKIMLEEGHTTAAQAE
SATYLGAUKLVDSLIGSLEQKLADLTIVKGNPLEDISNLKNIOHQVFKGGRVLPDQ
IH"
55939. .57297
/gene="BH2936"
CDS
55939. 57297
/gene="BH2936"
/note="BH2936
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06655.1"
/db_xref="GI:10175558"
/translation="MKGETKKQLTFALTWPFIETILLHMLMGNADTLMLSQSDDAV
AAGVSNQILAVIIMFGFVATGTSILVAOHLGAKERENAGKAVAVSIGANLIEGIVL
GLLIATFGPPTLKAMQLDSDLQEATLYLQIVGGFSVVQSLIMTAGAILRSHSFTKDV
MYVTIGNMLNVIIGNLYLFIGPFGIPVLGTVGVALSVWSRTIGLFWAIILLYKRIRG
ELPAYILKRPVRYELBNLLKIGIPSGEQLSYNASQLVITYFIAMGTREALITYKYIT
QNLMMFVFLRAVAGQGTQILIGHQVGAQIOAAIVRCFSLWTAMTVSVMVVFFA
FSTPLLGIFTNPDILLSLGTTLTLLTITILEPGRACNLVWISSLSLAAGDVKFPVYLAI
SMWGIAVPYIALLGLPLGLIGVWIAFIADWFRGLLMIWRWRKQWQESFVGRKA
SA"
57499. .58539
/gene="BH2937"
CDS
57499. .58539
/gene="BH2937"
/codon_start=1
/transl_table=11
/product="immunogenic protein"
/protein_id="BAB06656.1"
/db_xref="GI:10175559"
/translation="MKKKLGLSAGFLSLGILGACGGNADTDAEIPAGDDPADTGEE
QEEAQHLDLQLOLQSTGTYYPLOGEATANLVANVDYDGNFVAVASGASVDNIG
QIFRGENQLOTHVHPALQAIAGEGDEGAVVDNFGFLGYIYPEVMQVVTDEGTITS
IEDLKKRVAIGPAGSATQSAARLILBAHGIEDGDYEAFFEGFGDAARODGOLDAS
FGLLGPDSVTSELSLOREVVLPLEGDALAYVEENSDYGAYEIDAGAYDFLDDSVST
VTAYAILVASLDQVSEDLGHEIVKALYENAGGIITHPOGEHMKMENILNGSEGLPIHPG
AQYFQSQGISE"
58000. .60764
/gene="BH2938"
CDS
58000. .60764
/gene="BH2938"
/note="BH2938
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/protein_id="BAB06657.1"
/db_xref="GI:10175560"
/translation="MKEKQSEATVDKAEILRKRYKESNYRTNLQKWWVWVITILALS
LTLFQLYTSRGAYVTLIOGAIHLAGGLSIFILYPAKRSLKKKGVPWYDGVLLALIA
LAANFYIVFEYQRLTSNQVITLGVTDLDLIIVATGLVLLVEATRRVAGLPIVITATVA
LLYGLFGOFHPVQAQFRHAGFDEDRMLTLMFNTFAIFGTPRISSTFIPLFLEFGVNL
TKTGIGOFFNDLAFGATGRMTGGTAKAAVTASGLOCHWTCSSVANTVSSSFTIPMMK
RAGFKGFSAAEASASTGGOIMPPIMGAAAFIMASNTAIPYQEIILIAIIPALLYFT
GVFMGTFEAKGKIRGLPKELQPSLKGILRLDLAPLVAIIVMLYLGYSPTRAALF
GIGAAFFVLFERKDRMGIRIDIFDLEQGARIAPLVAACATAGIAGIVTTTGLGKG
IAGIIDLAGGVFLVIFTWMIACILLMGLPTTANYVVTVAAMAAPALMAFDSVPLIA
AHFFVEFYGVADITPPVCLAAVAGAGIARANPKAGVTAFKLAIAFIIPYAFMNP
VLLLYDANFVDIVIANSTALIGHAASVSNVFNVTNIIERILLFFAGVALLPQI
LSASGLVVLAIAGIQFRKRSQDQARATA"
complement(60805. .61542)
/gene="BH2939"
complement(60805. .61542)
/gene="BH2939"
/note="BH2939
unknown conserved protein in others"

/codon_start=1
/transl_table=11
/protein_id="BAB06658.1"
/db_xref="GI:10175561"
/translation="MMWELIFILFLLVGSFQGVSGFGFGLIAMSPLPFFETIKEST
LLVVALGVLTALFVLIQLFRIHRMKDLIVIMCAALIGRYVASFVLTAYGEMDLLKXSL
GVFLIIVGYLFDWKPQPEGRSVNOTWLPATLGGFGLVGVFAVGGPFFVFYFMLLF
RDDIAYASNIQVSFFITNFTTVIHGVHGDFTQSYLIYLIYLVGLISVFIGTRLGYVCF
TYLSQKIRIRLAGIVVAIALSLVFFS"
complement(61804. .62658)
/gene="BH2940"
complement(61804. .62658)
/gene="BH2940"
/note="BH2940
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06659.1"
/db_xref="GI:10175562"
/translation="MLNSDVYKRIMDLQEKMSQKKIANYLIQHPTAPFLTASKLA
KYAAVGATVIRFAVFLGYEGYDQLQRLHQLQALQKMTSAEVFARTTREDRPSHAIE
EILTDDMANLRATLQOILEFEQAVNSIIIAKRIYIAYRSAASLQGLFTFYLDLVL
QNTLVEQADGVSEHLLDITEDDLVIGGFARYTKRTVDVLKYVKQRCAKTIVTDHA
LSPLVPVIGDRVLLAATEINSFIDSAAPHISVSAFIATVTRSEHEKQVONRLOOLELW
ASFWNFI"
complement(62678. .64063)
/gene="BH2941"
complement(62678. .64063)
/gene="BH2941"
/codon_start=1
/transl_table=11
/product="glutamate-1-semialdehyde 2,1-aminomutase"
/protein_id="BAB06660.1"
/db_xref="GI:10175563"
/translation="MAWVEDTRPHTVAEOTPRSKSWIEQAKVMPGCGITANIKHPAF
YPIVKKCGCAFITDNDHQYVDLYLAYGALMLGHGHPYEQADENKPAKFEFGHHGG
PHPLEVTFGEIQOQLYPSMERLYTNSGTALLLHAMRIAQAYTNKRIAKFEFGHHGG
YNDVLYSVSPTLSEAGPADAPIVKESKGMHVTGDEEPLILPFNNLTACERLLRAQQD
SIAAMTEPELQGGFIPATDTFIAGLRELTIELGIVLIFDEYKTVFRLRGLGAQWYQV
SPDITLTIAGLKVDFVIGGRKDI MEVTSPTTQSDVFDQSTFRSRKQDVWHSYGT
YNGHPLILAAGLKTVDVLKDIKDVAFATDMLKDEALFVRKGVPMKAVGIGSIFSV
VLTEEDIFNVDLQASDLQTRKLLDLHLNNGIYTKPLNRYISLSTAHTEREVERTISA
YDRVLSRLFGG"
complement(64134. .64289)
/gene="BH2942"
complement(64134. .64289)
/gene="BH2942"
/note="BH2942
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06661.1"
/db_xref="GI:10175564"
/translation="MKKLNEKKPSLFSFGYVLGKERSTRGDOKKENLHLFLMKMLI
IQNWADT"
64382. .65563
/gene="BH2943"
64382. .65563
/gene="BH2943"
/note="BH2943
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06662.1"
/db_xref="GI:10175565"
/translation="MAQPIYIMNVDKIEQIPKEBEREKLQITDKFVRVNDYLYSLID
WNPNDRIRKLVIPNEGELSEYGRWDASDEDTNVVPGCQHKYETETALLICSEVCGY
CRYCFRRKRLFNDVKEAMSDVDPGLDYIAQTPOINNVLITGGDPLIATKRLRIIER
LRAIDHVKIIRIGSKLPVNPMTREDEOLLELLEIYREYSTPDHRIYIMAHINHPVEITN
EARAFOALHDAGAIYVNOTPVLKGINDDPDVLAELDLKLSWAGVTPYVPEINRPVAG
NNDEVLTLEEYNIYVEKAKAKTSGLGRVRLSHTSGKIEVLAIEDGKAYLKVKHQR
DGNTGKFWLDCPDATWFDLPGNENWTWKPEKKVKREVSVMEMPDPDPQKKRAATV
R"

gene

65710.. .66474
/gene="BH2944"
CDS
65710.. .66474
/gene="BH2944"
/note="BH2944
unknown"

/codon_start=1
/transl_table=11
/protein_id="BAB06663.1"
/db_xref="GI:10175566"
/translation="MKRLGGSTLLFLFLACGNTEFIIVQANNAHEPTDHRFEIG
THVVAEPGDFTEIEMLTSDPDVIDEDYLFAISHSIDESKEYEAOSSEKEHEVT
DQDQAVLLIRQFETPALESEAPSIDFTLYQPTYYEHEVFNKLDEHVSSEKNDLTI
RSDIDGKKLALVVEDVHNVKIELAKQOQEYIYVPSSITDHPHFNFSFHAIEYFAQ
SLPNTLTLMRRPHLDKIIWEFPLTATEIINTPAK"
complement(66506.. .68476)
/gene="BH2945"
complement(66506.. .68476)
/gene="BH2945"
/note="BH2945
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/protein_id="BAB06664.1"
/db_xref="GI:10175567"
/translation="MTEKQTIDEPKLSPBEQEILEKYDAESRPTFSGKKLPLLVSLL
AIGLSYHLTYSVFGPTLTLOHRSLVAVILALVFLYPPFKKANKSVPLXDWLCVL
LALSTIYFVDYAGIVHRGGIPNTYDLAGLVILVILEAARVTCMALPLLAVLFF
IYGLFQRELSQWRHRYGQWDELINMYVTGVGYTAIGVSATYIPLFILFGLSK
SGMQGFNDMALAIGQTKGGPAKVSIVASGFLSGINGAAIANVVTGSGFTIPLMKKI
GYNRFAGAVEAASVGGQILPPIMGAAAFINAEILGIPYSQIALAALLPALLFYAI
IYQILHRTAKLGLKIGSKENLPQVMDVLERGHLPLFLFLYMLFSGRTIYSASF
TIIVTVAMFRKTRMTIRDLIDALEOGARNAIGAVACACGIIIVGATVGTGFLK
LANGIVMLGESLFLTLEFTMIACIILGMLPSIPTYIITATMAAALVTIGLEPLVA
HLFVYFGLEPANTPPVALAFAAGISGDQMKTGFIEMKLAAGFIYPMFEVNEA
LILLNTTFLEGALVYVTSITGVVQMAAEGYLFITKPLFLRLLFGGALLFMNPNLV
QDLMGFALIGFVLLYQWWRGKQAVASFTVSL"
complement(68454.. .68942)
/gene="BH2946"
complement(68454.. .68942)
/gene="BH2946"
/note="BH2946
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06665.1"
/db_xref="GI:10175568"
/translation="MSRKAVFVSLCLFFLLAGSDLPSTLYVOGKNVYVYSTPVV
EGDRULTSIHVSVEKTPWEETIYIVKDGVMLEETVVFHSGAGVPHEKGTMTVEDGOI
VARGNQDQDALKWIIHSHSVHFQNLNGRVERFAGADLPHEKIEIIVIHBEVNEHDRKT
NN"

gene

65710.. .66474
/gene="BH2944"
CDS
65710.. .66474
/gene="BH2944"
/note="BH2944
unknown"

/codon_start=1
/transl_table=11
/protein_id="BAB06666.1"
/db_xref="GI:10175569"
/translation="MKRILIDDEPSTCSLSQFALEDDEYETEATTEVDKGLHIEQEP
IDVLLDKIGKVDGLEVLEQEIKRKPETIVITAYGTIANSVKAIIRLGAYSILTKP
IHMEELFSLVEQALHYKLNQEQVAYLSQELKKYEGYGMIGKSPAMTQVPHLIEKVKN
VDSVLIITGESGTGKELVARAIHFSGKRKQEHFEVNVCAAIPBHLLESELFGEKGA
TGAATKEGKFOALNBSGTIFLDEIGDMPALQAKLLRVLOKKEVTKLGSNKTLEKLV
VIAATNKSQEQKVEGTFREDLTFRLHVIIHLPPLRQERQEDLPYLQHLFLHTNQBL
GKSMKFRSDAQRELLAYQYGNIRELSNIVESSMVIADGEYIELEDLPMFIREKGG
FAKQKGSLODYIGRPLKELEKDFILATLKHNDNHRKKTAEMLGTSERGLRDKLKQYLV
GKN"
complement(71517.. .73520)
/gene="BH2949"
complement(71517.. .73520)
/gene="BH2949"
/codon_start=1
/transl_table=11
/product="two-component sensor histidine kinase"
/protein_id="BAB06668.1"
/db_xref="GI:10175571"
/translation="MRRFPLYLMKLVFMFVMLPAPALAEQVKVAYDPHLPPLHAGG
DDPGYGVIDIFSEVAERKGDHAYVPMTORQAIETADQKVDIFLSLPKESYANRL
EFTDSILGTSGILIVETNNEEQIQLPDLANKLSHKYRVAGNYLLPDYTTAVOKENV
QEAIEILLONDRAEAFITGSVLTANYLLLEOKLGLSHKRVAGNYLLPDYTTAVOKENV
LMDINDGLRELXSDGTYSTLYQWENGEOQVTERLWLAQIIIGVLLIALLILFFLGI
RWRQLOKEVAKTAVVHELNRLEKQIEETNRNSFORQILNKPSPGATVNRREGI
TSPNKAATVGTTRSVGLHFGKLSKLLNLLIGTQVTFEORFLGEETKWERADG
RTIYLYYVYFVDEGNVTGTTIVTIEERQRMQIFEQEKSRALIRVVAGIAHE
VRNPLSITFVELIPKLCNERFOKEISTYVPQETIERVSTLIEGLIDYAKPKQPSRE
MIDLAELARECMLLDRVTNRGTMDGATEGFLVKADPNOLKQVVMNLMINSLOAM
KENEGRLSIRVYQLOKQEYILEVDGEGVTEQEQKALEPFTTKAGTGLGLAI
ANQVANGGRVIRITARIGTITSLHFFMIREGTTDETNSHN"
complement(73616.. .73684)
/gene="BH2950"
complement(73616.. .73684)
/gene="BH2950"
/note="BH2950
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06669.1"
/db_xref="GI:10175572"
/translation="MKHRVLLFSIEKKGKGLINTFN"
complement(73700.. .74272)
/gene="BH2951"
complement(73700.. .74272)
/gene="BH2951"
/note="BH2951
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06670.1"
/db_xref="GI:10175573"
/translation="MTIIMPTLWMVSVILFTIGVSFLLSVIMSKFLKKGDKKENNA
GFLAAVVTLLIVATSEAPVERVEGDVLTADKANEVEDATARYLSTERPLVAVNF
RHGYIESIETNELGTWRHMQVTHYPERAMELHSEGHFFFPQYIEAVVQPVVT
DVLEESPPLSASAMEEMINDRLPGHEVNLNQ"
74361.. .76139
/gene="BH2952"
74361.. .76139
/gene="BH2952"
/codon_start=1
/transl_table=11
/product="Mg-protoporphyrin IX monomethyl ester oxidative
cyclase"
/protein_id="BAB06671.1"
/db_xref="GI:10175574"
/translation="MKTVVTTLNAKYIHTCLARCLKAYAEPEPPIEMVEYTIKDPAM
NIVADLHQRPDVIGFSQIWNQETIQVIEMLKVMPSVKIYVGGPEYSYDTREWME
RLHAVDFIVVGEGETFKHLLSEEMNTAKVHLVFGLAYRKGGDIVINPPRPKLPLNEL

gene

65710.. .66474
/gene="BH2944"
CDS
65710.. .66474
/gene="BH2944"
/note="BH2944
unknown"

/codon_start=1
/transl_table=11
/protein_id="BAB06665.1"
/db_xref="GI:10175568"
/translation="MSRKAVFVSLCLFFLLAGSDLPSTLYVOGKNVYVYSTPVV
EGDRULTSIHVSVEKTPWEETIYIVKDGVMLEETVVFHSGAGVPHEKGTMTVEDGOI
VARGNQDQDALKWIIHSHSVHFQNLNGRVERFAGADLPHEKIEIIVIHBEVNEHDRKT
NN"

gene

65710.. .66474
/gene="BH2944"
CDS
65710.. .66474
/gene="BH2944"
/note="BH2944
unknown"

/codon_start=1
/transl_table=11
/product="immunogenic protein"
/protein_id="BAB06666.1"
/db_xref="GI:10175569"
/translation="MKRFLYLMVVALFCFSILVACNGEGASTDBGTANGGGGTAPIDD
LFVTVASGTSYVPIAGIAQIILESDIGLATSVQATGASVENVNLIINSRAELAIT
MADTVLAQYEGEGFEQEP IETLRGLASLYPNFVOIVTTEDSGIETVDDLHGKSVGV
GAPNSVLEARNALVLEAHGMSYDIIEEDLYSSEAIDQIKNGLIDAAFYTSQVNPATV
IDLGTHSAKIPIEGQANDYLTENYPPFSEGAIPACSDNDEEIEPTAAITNVLLISH
EIDEEVAYETKLSFELHDEIHASHNAATDISLETVEVGMPVPFPHGPAERYKEVGVL
D"
complement(70178.. .71539)
/gene="atoC"
complement(70178.. .71539)
/gene="atoC"
/note="BH2948"
/codon_start=1

gene

65710.. .66474
/gene="BH2944"
CDS
65710.. .66474
/gene="BH2944"
/note="BH2944
unknown"

/transl_table=11
/product="two-component response regulator in acetoacetate
metabolism"
/protein_id="BAB06667.1"
/db_xref="GI:10175570"
/translation="MKRILIDDEPSTCSLSQFALEDDEYETEATTEVDKGLHIEQEP
IDVLLDKIGKVDGLEVLEQEIKRKPETIVITAYGTIANSVKAIIRLGAYSILTKP
IHMEELFSLVEQALHYKLNQEQVAYLSQELKKYEGYGMIGKSPAMTQVPHLIEKVKN
VDSVLIITGESGTGKELVARAIHFSGKRKQEHFEVNVCAAIPBHLLESELFGEKGA
TGAATKEGKFOALNBSGTIFLDEIGDMPALQAKLLRVLOKKEVTKLGSNKTLEKLV
VIAATNKSQEQKVEGTFREDLTFRLHVIIHLPPLRQERQEDLPYLQHLFLHTNQBL
GKSMKFRSDAQRELLAYQYGNIRELSNIVESSMVIADGEYIELEDLPMFIREKGG
FAKQKGSLODYIGRPLKELEKDFILATLKHNDNHRKKTAEMLGTSERGLRDKLKQYLV
GKN"
complement(71517.. .73520)
/gene="BH2949"
complement(71517.. .73520)
/gene="BH2949"
/codon_start=1
/transl_table=11
/product="two-component sensor histidine kinase"
/protein_id="BAB06668.1"
/db_xref="GI:10175571"
/translation="MRRFPLYLMKLVFMFVMLPAPALAEQVKVAYDPHLPPLHAGG
DDPGYGVIDIFSEVAERKGDHAYVPMTORQAIETADQKVDIFLSLPKESYANRL
EFTDSILGTSGILIVETNNEEQIQLPDLANKLSHKYRVAGNYLLPDYTTAVOKENV
QEAIEILLONDRAEAFITGSVLTANYLLLEOKLGLSHKRVAGNYLLPDYTTAVOKENV
LMDINDGLRELXSDGTYSTLYQWENGEOQVTERLWLAQIIIGVLLIALLILFFLGI
RWRQLOKEVAKTAVVHELNRLEKQIEETNRNSFORQILNKPSPGATVNRREGI
TSPNKAATVGTTRSVGLHFGKLSKLLNLLIGTQVTFEORFLGEETKWERADG
RTIYLYYVYFVDEGNVTGTTIVTIEERQRMQIFEQEKSRALIRVVAGIAHE
VRNPLSITFVELIPKLCNERFOKEISTYVPQETIERVSTLIEGLIDYAKPKQPSRE
MIDLAELARECMLLDRVTNRGTMDGATEGFLVKADPNOLKQVVMNLMINSLOAM
KENEGRLSIRVYQLOKQEYILEVDGEGVTEQEQKALEPFTTKAGTGLGLAI
ANQVANGGRVIRITARIGTITSLHFFMIREGTTDETNSHN"
complement(73616.. .73684)
/gene="BH2950"
complement(73616.. .73684)
/gene="BH2950"
/note="BH2950
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06669.1"
/db_xref="GI:10175572"
/translation="MKHRVLLFSIEKKGKGLINTFN"
complement(73700.. .74272)
/gene="BH2951"
complement(73700.. .74272)
/gene="BH2951"
/note="BH2951
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06670.1"
/db_xref="GI:10175573"
/translation="MTIIMPTLWMVSVILFTIGVSFLLSVIMSKFLKKGDKKENNA
GFLAAVVTLLIVATSEAPVERVEGDVLTADKANEVEDATARYLSTERPLVAVNF
RHGYIESIETNELGTWRHMQVTHYPERAMELHSEGHFFFPQYIEAVVQPVVT
DVLEESPPLSASAMEEMINDRLPGHEVNLNQ"
74361.. .76139
/gene="BH2952"
74361.. .76139
/gene="BH2952"
/codon_start=1
/transl_table=11
/product="Mg-protoporphyrin IX monomethyl ester oxidative
cyclase"
/protein_id="BAB06671.1"
/db_xref="GI:10175574"
/translation="MKTVVTTLNAKYIHTCLARCLKAYAEPEPPIEMVEYTIKDPAM
NIVADLHQRPDVIGFSQIWNQETIQVIEMLKVMPSVKIYVGGPEYSYDTREWME
RLHAVDFIVVGEGETFKHLLSEEMNTAKVHLVFGLAYRKGGDIVINPPRPKLPLNEL

PSPYRFSEDAVSLKRVITYETSRGCPYSQCFCLSSIEVGVYRDLERVKADLLYLIN
 SGARKLIFVORTFNIRDYALEVFEFLIENHOGCVFOFEITADIMRPEVLAFLNEHAP
 KGFREFIGVQSTNDLNLNELVORNEKELKRTVTVMKEGKIDQHLDLIAGLEEDY
 QSFRTKNDVPMHDELOLGEFKMLKGTGLRAEDHNYLFOQAPETILGNVLSF
 DEIVRIKVEDVLEKYNNAHRMDTILYLVYEYEDSPDFQSGDYGDEKGMARGH
 QLEDLFLRLFAFLEDRDNLPIIKGLMVDYFRGHKPKRKTWPPSFLTKEQSSLI
 KRLAATPEVSEALARHQLSEKELHKHMTLVEIPFSLADWLKRGVIVKEDTLLLSLQ
 PKGTEVFGPLHSLTNQKEETIHE"
 76132..76887
 /gene="BH2953"
 76132..76887
 /gene="BH2953"
 /codon_start=1
 /transl_table=11
 /product="NADPH-flavin oxidoreductase"
 /protein_id="BAB06672.1"
 /db_xref="GI:10175575"
 QVSVIVIREPNERELFAELVGGQAHVAQAPFLVFCADFHRAVAHAARCKFYLAALQD
 VDALMGVATDGIATLSNAITAAESLIGTVPIGGIRBQPLRVLETLOLPYVIPISGL
 CIGYPETVQKPLPLDVTDHGRYEPFLSQKLAAYDELISGNTKORTGGATTIWS
 DRLISFEQYQYQEIIVAMLKQGGFTCKNKEEG"
 complement(76898..77920)
 /gene="BH2954"
 complement(76898..77920)
 /gene="BH2954"
 /note="BH2954
 unknown conserved protein in others"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06673.1"
 /db_xref="GI:10175576"
 /translation="MEPPKKEKLTVEDYINGVLADGRALLARAITLIESHSEHRPITA
 TEVLNGILPOTGKSRIGITGVPGAGKSTLIEALGTMLCDKGRVAVLAVDPSSQLTK
 GSIGDKTRMEELSRNKNAVYRPSPSGTLGGVTRRSRETLTLLCEAGDFVIVETVG
 VQGSSEVRSVDCFLVMTLGTAGDELQIKGVMEEMADAIFINKADGPNRQVATK
 AEXNLLHFLPOATEGWETKAYTVSALTGBGISEMMDVIVSFFQFSKENGIFERREK
 QALEWYFAYIEDELKWFQODDIKQOLPLLMQMRKGEVVPYVAASQLLTFTFNGLE
 KRKEEG"
 complement(77982..80132)
 /gene="mutA"
 complement(77982..80132)
 /gene="mutA"
 /EC_number="5.4.99.2"
 /note="BH2955"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06674.1"
 /db_xref="GI:10175577"
 /translation="MQTPDFATMSPPRSLSTKSKEANVAFETNEQIPVKTYFKEDI
 HGKKHUNYVAGIPPPFRGPYPTMYKTRPWTIRQYAGFSTAESNAPYRRNLAQGKGL
 STAFDLATHRGYDSDRHPRYGVGVGRAGVAVDSILDMKILFDGIPLDQMSVMTMGAI
 LPVMAFYIVAAREQGVKQQLTQNDILKEVMYRNTVLYYPEASMKYADIADFEVTS
 NYMPFNSTISISGYHMOEAGATADLEAVTLADGLEVYFTGMKAGLIDIDAFAPRLSFF
 WAIGMNHFEVAKMRAARLIWAKLKRPNKPKSLAKTHSQTSGWSLQETQPNVNV
 VITCVALAALCHTQSLTNADLEAIDPTFSARIANTQYILODETGITNVLDPW
 GGSYVESLTAELLQKAWANIEIEELGMAKIDTGLPKMRIEAAARRQAKIDSGK
 EAILIGNLYRPTDEPLDILDINTVAREAQIRLEKKAERDEERVOQALNRITKAA
 ETEKENLLTCVAEARASALSGEISMAVEKVVGRHKAVIRISGVYRSEFGEKEEAR
 VREMTDITFYEREGRRPRIWAKMGQDGHDRGAKVATADMGEDVDJGLPLQTPPEA
 ALQAVENDVHVLGMSLAAGHKTLIPQVDELKKLGRDINVLIGVIPPQDYDFLKA
 HGAAAIVFGGTIIPCAQOQVLMHMRRLYEGEA"
 complement(80119..82164)
 /gene="mutB"
 complement(80119..82164)
 /gene="mutB"
 /EC_number="5.4.99.2"
 /note="BH2956"
 /codon_start=1
 /transl_table=11
 /product="methylmalonyl-CoA mutase beta subunit"
 /protein_id="BAB06675.1"

/db_xref="GI:10175578"
 /translation="MTQVLDLDFAEPIPTPYDEMRKVAEASLKGAAPDQKLLTKTLE
 GLTLEPMYQORDDELHISERHVPKRNKGLVWSQELTQCTPESFNRTAKNLIARG
 QTAJHMVIDOTVRDGLSPADQGETGRIGTLLYNHEDMETALEGDFSSVPIHLNGGA
 VQTFPLASLLAIAPTSLHGVVAGDP IHEWVKKGLSYSWRTVDFELSHSVOVNERAL
 DLKTLVQSHVYHNGGATATEELAFLLATGVQVYRALLERDIAIDDIGRSCMSVSVG
 SDLEKELAKRALRLWANIMEAGKEGRNIMLHARTSFYTKHITVDVYVNMILKGT
 EAFNAAVGGDSLHVSFPDEPIQKPSFRIARNTSIIIQEEAHIGQVEDPAAAGSWY
 VEVLTEKLGAAWSLFTQIESEGMIASLEKGLPOOWVERSROEQAAEVKRTVMVG
 TNRYVAIDEPKPLTQEEEPITAPYREARTKQEKHLKASNYPSLLQAVKEGVPPFKLLT
 ALQTKEAAPTIASIKTSLMSYEALRKATBQWVKRTGDRPRVTLTLTGLSLASHKPR
 DEARELFQVGSFOVDTTTGTSTDEALRQLQEK"GNPVLGCRDEDESLAASLIEAV
 REAHPNAMILLAGNLSSETQAYEQKGLYGTIHTVNTVQFMNEVLMKRGVIRDANT"
 complement(82217..83767)
 /gene="mmdA"
 complement(82217..83767)
 /gene="mmdA"
 /EC_number="6.4.1.3"
 /note="BH2957"
 /codon_start=1
 /transl_table=11
 /product="methylmalonyl-CoA decarboxylase alpha subunit"
 /protein_id="BAB06676.1"
 /db_xref="GI:10175579"
 /translation="MDMFDRIHEMEVRRDKVKGGSKRIDAQHDGKGLTARERIDLL
 LDGTEVELNPPFIEHRGLDFGTAEAPGEGVVTGYGVDGRLVFLFAQDFTVFGGALGE
 MHARKITKVMDLAAENGAPIVGLNDGGARIQEGVSLDGYGEIFYRNSIYSGVIPQI
 SVILGPCAGAVYSPAITDFVFMVEKTSOMFITGPKVIESVTGAKINSEDLGGAKVHS
 TISGNAHFTAAEEGVLQVNRLLSYLSPSFEKPHQRAHPVPLDERVOQLLIDV
 VDCTKYVDVHVIEKVVDDEQFLEVOQPQFARNIYVVGPARINGDTVGLIANNPKMAG
 LDINSDDKCARPIRFDCFNILPITFEDVSGFIPGVQOEHGGIIRHGAKILYAISEAT
 VPATIVILKRAYGAVVALNSKAIGADLVFAMPNPAELAVMGPEGAANIIFAKEIKESA
 NDRATQAKTISEYRERFANFYAAANGMIDVDIPDRTRKSLVSALAMLNKKNKTLPK
 KKHGNTPL"
 complement(83771..84418)
 /gene="BH2958"
 complement(83771..84418)
 /gene="BH2958"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator (TetrA/AcrR family)"
 /protein_id="BAB06677.1"
 /db_xref="GI:10175580"
 /translation="MKRSTKERIETALLFEBHGPHAVTVDRIVQESGTSKGFYHN
 FKSKDELYTIHDSFTYVLEKAREAYEKYETPAERLHATVRSFVMFDFYRPHVTVF
 QYESLYSSDYFKDIKKRQDYKQLMFRLVKEGIDAGEPREETPVPTTSMATFGMINW
 TYKWKYKSGMYSVAETADIVADLVHLAVLTQREALTRPDYQSLFLQHLGRIGERSV"
 84568..84853
 /gene="BH2959"
 84568..84853
 /gene="BH2959"
 /note="BH2959
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06678.1"
 /db_xref="GI:10175581"
 /translation="MCGVVVKFEKAKMQSLEHDFRLRETYNDTIQVWDEEALRLIYD
 VMILKEPQONAYLHLT"
 complement(84922..85215)
 /gene="BH2960"
 complement(84922..85215)
 /gene="BH2960"
 /note="BH2960
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06679.1"
 /db_xref="GI:10175582"
 /translation="MSOHFAMERQALROLRLKRVKRPESVLQSLVAQAILEVAFYAF
 RKKEIDEAIDQALEKRDQDAPMSLTNKRQVFIDEYKNGKIVEEQGHKLQLLFF"
 complement(85509..86285)
 /gene="BH2961"

CDS

complement(85509. .86285)
 /gene="BH2961"
 /note="BH2961"
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06680.1"
 /db_xref="GI:10175583"
 /translation="MMDEIEIOOLKMEFFVHTFPVEKVDLMTITGLVCKEGAWKWE
 WRNGILKKEKLVGSGAMIKPESLLFAAHYVLASVKNOMLLFDLAFPSKCKEUGFEL
 RRARLEPLALNDISSYFTALAKRIAPIIENTAEVADLKTAFAGLLAPLPYQQQSAWL
 KRETQDTLELRITDGEVLKQSGDAFLGRNPFDFVKFYSSWADPNHVRIRKPACC
 LSYLKEGNYCYSCPMDTEQREVRAAQLREQAAQQKSS"
 complement(86289. .86510)
 /gene="BH2962"
 complement(86289. .86510)
 /gene="BH2962"
 /note="BH2962"
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06681.1"
 /db_xref="GI:10175584"
 /translation="MRQLEIIQRWIDENOLVVISLLEDORKVIGRILHYDSKNRACL
 YDDGKSVHHLALTEIQSIKRTDGMKKG"

gene

86630. .86989

CDS

/gene="BH2963"
 86630. .86989
 /gene="BH2963"
 /note="BH2963"
 unknown conserved protein in B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06682.1"
 /db_xref="GI:10175585"
 /translation="MLDEVRNLQVRYIILTVRQALEKDFNTLQASSLFCQESLHSF
 LTERHVEGREIGALORALKQKLTKRKNDFGFTKSYYPKGEGFTSLNARLKH
 DALTYMKRHLFKLETKQ"

gene

complement(87369. .88202)

CDS

/gene="BH2964"
 complement(87369. .88202)
 /gene="BH2964"
 /note="BH2964"
 unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06683.1"
 /db_xref="GI:10175586"
 /translation="MNKLFDFILTGSIIVAVGLELILAPNSLVDDGGVTAIAMLNA
 IFNLPIWLVLLNIPSLYVSGYMGKKFIIRTYANIITAFALNMAPPFAITTSV
 LIVLKGGLLVGVGLVKGAGAIIDGTEMLAIWINKYHIPISKFLLLYNALILSTAA
 FVSLGLOAMFSTAVFFIVTKTIDFILDGINQCKSVMIISEKPEEIQOQLIDELGMTIT
 YLYGEGGYSNQEVHLLYICITDRFPYKKEIALSDPSAILEASFVTTTGTGVRKSLF
 N"

gene

89272. .90444

CDS

/gene="BH2965"
 89272. .90444
 /gene="BH2965"
 /codon_start=1
 /transl_table=11
 /product="transposase (04)"
 /protein_id="BAB06684.1"
 /db_xref="GI:10175587"
 /translation="MSTSIGSTFENOLDMSVREFVKEKLETIMKEMESFTTHEUPE
 LKNQNGFTROLDTRYGRLENLRVPRDRQNAFQTELSFPYQRRRENLEALITMFK
 GYSTREIGHTIEKILGHYSASTISQVTDVVTETQWQORPLKRRYSVLYLDGYLK
 LREADVANEVYLVVGTVEEGYREILGYVGQGESANGRNILDLISRGLEEVILG
 FQGLAGLEAMKAVYPKADVQRCVVKRNALHAVRKKDQPAVADFPIYQANTYEE
 AKNRKFEKENQKRYPKVVAWTEKDLVLLTFLSPSTIRPMIYTNIIETMKEIK
 KRTKTMNSLPNEKAEKIVYLQADIDYNERWATRLRGFGAEYALQEMFKQRYSGP"

gene

90607. .90732

CDS

/gene="BH2966"
 90607. .90732

gene

/gene="BH2966"
 /note="BH2966"
 unknown"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06685.1"
 /db_xref="GI:10175588"
 /translation="MKQLIKKLFSHESIRGLIIGVGMIFVFLFVHYILGIKAFK"

CDS

91478. .92803
 /gene="BH2967"
 91478. .92803
 /gene="BH2967"
 /codon_start=1
 /transl_table=11
 /product="transposase (01)"
 /protein_id="BAB06686.1"
 /db_xref="GI:10175589"
 /translation="MLGFLHKGQKKCTQAPLSIIELSRNKRIGCACTMMFTMDLP
 GLKGVKVTMEEREGFIALHVMEMPTPHRCPCSGERTERTHYRIQKVQHLKLFERW
 YLFYKRRYACRGKRFKACFVERYQRTVEMNQAFRLRIQGNFTDTANQFTS
 VTTAMRRFDOLAAPLVKVDHLPPIAIDDYKGDYNAGKYQVIADAPTRKLDILPD
 RKADTLKRYLOEKGHVKTVMVMDLSYTFKSAVQKALGHPLIIGDPHFRCRYIYWALEG
 VRRRVQKAFHPYHRKRCRKRIKHLFKOKEREGLDKOKKKVRYLESLSPDLREAYELKEH
 FRQYQEARMQGREGRMSEVKQGLYAFYRVEASPLKEPHRAIQTLKNQQTILLNSFA
 FGLNNGFTLEGLNNQTKVIKRNAGFRPRYDRLSRILLHHQFKHQTFGVG"

gene

complement(92995. .93357)

CDS

/gene="BH2968"
 complement(92995. .93357)
 /gene="BH2968"
 /codon_start=1
 /transl_table=11
 /product="two-component response regulator involved in
 modulation of flagellar"
 /protein_id="BAB06687.1"
 /db_xref="GI:10175590"
 /translation="MATVLVDDDAAFMRMMIKDITKDNFVVGAEANGEAEAVQLYKE
 HSPDLTLDITMPKDGIOALKEIKAYDADAKVIMCSAMGOQAMVIDAIOGAKDFIV
 KPFQGRVIEAINKVLGK"

gene

complement(93381. .93848)

CDS

/gene="cheW"
 complement(93381. .93848)
 /gene="cheW"
 /note="BH2969"
 /codon_start=1
 /transl_table=11
 /product="modulation of CheA activity in response to
 attractants (chemotaxis protein)"
 /protein_id="BAB06688.1"
 /db_xref="GI:10175591"
 /translation="MVSSEMATKNDLKLIVFQGDDEYAIEVDYIQSIERMQPVTRIP
 GTPSVVIGVNLRGVITPVINLARFGLPAKADDDSTRILVIADDDGVEIGFVVDGAND
 VVDVPAEKTEPTPEVVGQEDYLRGVVKEEDRLFTLLNLEKVIQRKKQLST"

gene

complement(93864. .95912)

CDS

/gene="cheA"
 complement(93864. .95912)
 /gene="cheA"
 /note="BH2970"
 /codon_start=1
 /transl_table=11
 /product="two-component sensor histidine kinase involved
 in chemotaxis"
 /protein_id="BAB06689.1"
 /db_xref="GI:10175592"
 /translation="KSHSEYLFVIDESOEHLQAINDLNLLKLEVSPSLISYVCEPFRS
 AHTLKGMAATMGFEDELQHTHNENVLDLIRNQKLDATSEVLDPVEAVDDLEAMVGD
 IASGGKRDVTKVVMLEQIEKKGTPDAGVKTQVTPAAGQLELSGSSIIIESYDGE
 MVLQDSFQQQCAQYIEVTLDEKTLKKAARFMFVEVLEQVGEVTKSTPSAELEEE
 KDFERFLVTLTKDGEVVRQRLIKVSEITAVDICSLDKQALQALKQTOHQSDKQV
 VKAEQTEKKTATTPSDKKGVEVNTYRVNIRLDVLMNLFEEVLIDRGLRQVLS
 ELNQNELTETVRMSRISGDLQEIILNMAMPVEQVFNFRPMVRSLSRELNNKYNLE
 IYGAETDELTRTIDEIGDPLVHLIRNSIDHETPEVRAAKGPEEGTVILRAFHSNG
 NVFIETDDGAGIDRDKVLQKALTNGVTEEDAAMNDQVQLLPAFGFSTAEKITD
 ISGRGVLDVWQSTFESIGGVSVTSKLGKGSLSIQLPLTLISIIDVMLVEVADEKYA

```

IPLSSIVETAIVNKNVEYSAHNKVIDFRGKVVPLVFLSDIFAVPCKREADELYSLVI
VHKGEKVAGLVVDLSLIGQHDIVLKSGLDYLTDVFAISGATILGNQVVALIVDYNALIK
"
complement(96072. .97703) .
/gene="BH2971"
complement(96072. .97703)
/gene="BH2971"
/notes="BH2971
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06690.1"
/db_xref="GI:10175593"
/translation="MPNQSNESRPFHKDLBSILYAHVPLPMFTMDREGALLNNNEGAM
ETLVRDREQNORSVQOLFSECRESLYLALATVNNKISQVEKVELIKRNGETETFCL
SRLQLDPKQVIETIADIDKMDPLTGVPRIQVPELMTFQENQAETVFLFLINLD
RFLKLVNDSLEGDQVLOTLVRRLLWQLPNQOWIARVSGDEFIACAAIKGELPKKET
LATYISSIIQQPFQVDEVOFTMTASGAAYPMDGKSVEMWNLSERALRHAKARGN
CVSWFOELKDEWIHPERLENDIRRAVENOELEYFQPIFETTTGHCCEEVLYVRNH
PALGVLGPNFFIPVAEEIGLIGVGNWLDACKQKALLIEKASVPMVSNLSLQPF
LQPDIVMDILDSUNKIGLDFSLSEVEYTERITIDMRAMKVLOKLRQAGIRVSLDDFG
TGYSLOYISQLPIQKLIDKSFIAEIGVNSQVEATISMIHLGHIELDLTVAEGVET
EKOLSFLKEKCDTVQYYHAKPLTFSEFKKYLSS"
complement(97742. .98572)
/gene="BH2972"
complement(97742. .98572)
/gene="BH2972"
/notes="BH2972
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06691.1"
/db_xref="GI:10175594"
/translation="MNPVLKSKLVIFFDLGLTYEGTDHFDYYANHKKQNPQSQQLD
FDPVSTAMDKDHPHVGIGKYDYKDYILTIDPMTISVQSVQVTDGYSYVPAEWEETLY
SRVPSDFETIITAVGDGWNYPFVTAHYGVDCYASYVATKEVMYDFEFLPLPGLR
EGLIELKATKEVLVNTSDREDVQRLHELHLDVFNHIIITSAKPTRTDIFKEVVH
QYEVDMSQVSVGDNFMNEIAQALTLGMQGIYIHQRASEVAHDHLTIVRSVVELFKN
"
98831. .99772
/gene="phnD"
98831. .99772
/gene="phnD"
/notes="BH2973"
/codon_start=1
/transl_table=11
/product="alkylphosphonate ABC transporter (phosphate-binding protein)"
/protein_id="BAB06692.1"
/db_xref="GI:10175595"
/translation="MKKMLYFLGVSMVFLVLAACGTSDDGDTAGSGADGADGAEAGEMP
SELIMGFPVSSSDQIADTVEPLAERLSEELGIPVSGQVMTNVTGLVEAMGNDQVHIG
FIPAGYVATRDYPOVEVILKSVRNGSTYRAQYTVLADSDIESIEDLEGKVVYISD
LASTSGYLPFAQLVTEYDVPADGIDSPFGNIOAGSHDNALITVLEGADVATTFED
AREIIEGYPEAMEDLRIGYTEIPNDTISVLNTPLOELIDQIRDVFLSFNEDDMI
KILNDVYRWDAITEASDDEYDIVRETYELFSDQVSLD"
98892. .100668
/gene="phnC"
98892. .100668
/gene="phnC"
/notes="BH2974"
/codon_start=1
/transl_table=11
/product="alkylphosphonate ABC transporter (ATP-binding protein)"
/protein_id="BAB06693.1"
/db_xref="GI:10175596"
/translation="MIEFKNVLVYPNGTOGLKDVNLKINEGEFVYIVGLSGAGKSTL
IRSNRLVTPPEGLLIDGENILYPSRQLRKLRTKVGMIIFQNYLNRSSRVNKRVIS
GRGHVGTFSALLNLYPKREDVALAYRSIQRVNIAEKYQORADQLSGGQQQVAVARVL
TQQPKIILADPEFVSLDPPTSHQWMTYLRKINKEDKITIVNLHPIDNAMEYADRIIG
MRAGEVDFGPASDVSEETFEIYGRKIREDDLVGGQDNE"
100661. .101452

```

```

/gene="BH2975"
100661. .101452
/gene="BH2975"
/codon_start=1
/transl_table=11
/product="alkylphosphonate ABC transporter (permease)"
/protein_id="BAB06694.1"
/db_xref="GI:10175597"
/translation="MSKQEVKGVISSKAKLRMTVLLVAIVGMVFTSARTGSSIPDLI
SGSPRISSEMGELFPPNMGVYASOVWGLAFTVQMAIATTVAAITIPISLISAANLF
PNRTIHPVRFIMNVLRTIPDLLAVLFGVIGIGVPSGIMALLIFSLGLLAKLLIYET
IEAIDLPLEAIRASGSGNTIQVIWAVMPQVLFQVFSLYFEVNVNRSVVLGFGVGA
GGGILIQOQLGFFNYPNVSTIIIIIPAAVVVVDIYISNKIREGLV"
101452. .102252
/gene="phnE"
101452. .102252
/gene="phnE"
/notes="BH2976"
/codon_start=1
/transl_table=11
/product="alkylphosphonate ABC transporter (permease)"
/protein_id="BAB06695.1"
/db_xref="GI:10175598"
/translation="MNLTLPPKKPRSAKTIRNIAVIAVIAIYVTTFTTINIDWGRI
FSERTIENMGRIIPQLFSPDMLYISKVSAMMETLAMAYGTLMAAIILAIPFGFAAS
NMVKRFFNTLSKWLDAIRAFPELMAIIFIIVAAVGPSPFAGVLAIGAIGMLKLY
CEVYIDMNVLEALGANKLOVLVYGIIVPOLPEFLSYAIVRFEIDVRASTILGI
VGAGGIGMLVSIATONRNNDVGMILLVIIIVVTLDIYISAIRKRIV"
102331. .102390
/gene="BH2977"
102331. .102390
/gene="BH2977"
/notes="BH2977
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06696.1"
/db_xref="GI:10175599"
/translation="MIYIVKGSIIFFVGKKQI"
102473. .103687
/gene="BH2978"
102473. .103687
/gene="BH2978"
/codon_start=1
/transl_table=11
/product="transposase (02)"
/protein_id="BAB06697.1"
/db_xref="GI:10175600"
/translation="WNFTMNFGLKDTFTTKVEEVEIIVLVYEMERKVRHPCPKCKR
TKKHIDYRIQIKHLKWERQTOIFYKRRRYVCACGRFAENNSFVRYQRTYSIENWQ
AVAIRAIGKTEKTAENYSSATTVIRFDQISSEIRPVEELPAIIAIDYKIGDTK
EGYQLIITADATYKPEPLDILPNRYKTIDYLRHRGSKVQVIMDMSPSFKAQVQAL
GRPVIIDRHFRCRVLYWALDGVRRRVQKDFHAYDRKKCKRMKHVPHFKNSLTERER
WDLVDSYFENMTELOKAYELKEAYQDNFRKSKEIGATQIERVKKELKAIYKLVKNSGIP
ENCRAIGTFENMQTEILNSFAYISNGPLEGINNSTKVLKRNAPGFRFSYERFRAKILL
THYKRRIGVHIG"
103857. .104156
/gene="BH2979"
103857. .104156
/gene="BH2979"
/notes="BH2979
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06698.1"
/db_xref="GI:10175601"
/translation="MKGOSVHIGEOYSQPEPWEHWYFCQTADGSLKGVVPKOLLAIEG
SKGVALDDYTATFELNVNDGVEIIVLKLKELNGTWICRQQTSGEYGVWIPKENVEKRVK"
complement(104402. .107791)
/gene="BH2980"
complement(104402. .107791)
/gene="BH2980"
/codon_start=1

```



```

CDS
/gene="BH2989"
114740. .115186
/gene="BH2989"
/feature="BH2989"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06708.1"
/db_xref="GI:10175611"
/translation="MKMKLKSFVLLGSCFFISMNOSISYATPSIIDFELEMLVKDY
STYDIEMENSLVDIATVPNEPTIKGQAEVKEVQSFIGTLQLLPSTSMQAVQNVL
TQQLSKDDIHFFKLEATFDGATLNIHWAGKCCNDCQHNERD"
complement(115257. .115907)
/gene="BH2990"
complement(115257. .115907)
/gene="BH2990"
/codon_start=1
/transl_table=11
/product="transcriptional regulator involved in phosphate
transport system"
/protein_id="BAB06709.1"
/db_xref="GI:10175612"
/translation="MSMRKFHSQDLYKQGIILMGHGHQVETVIEEMVAFQTFDLERMD
SIIEGDVRIHQLELDHQVTLISQQPVATDLRLKILVAKISGDLERVADLVWDIA
KQSSRLKSDSLPDYKQOLIDFQIAQSMIOQAVAYEKSDVLAQKALALDDKVDWTMY
GQPIRELFALPRESIEIDQVMAFVGRYLERIADYATNVSEWVYEVNKGKFDLN"
complement(115922. .116758)
/gene="BH2991"
complement(115922. .116758)
/gene="BH2991"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter (ATP-binding protein)"
/protein_id="BAB06710.1"
/db_xref="GI:10175613"
/translation="MALTVEKKKNVEVVPQPRKHVIOGQETVAYETRLNLWYKGD
HALKNTLSIYEKEVTAIIGSCGKSTYLKTLNRMWELVPSVRISGNISYGRNILD
KSFQVEELTRVGMVFKPNPKSIFDNVAYGPRHIGRNKILSEIIVERSLRGAAL
WDEVKRLHENAYGLSGOQORLCIARCLAEIPDVLMBEPTSAIDPKSTLKIEELIQ
ELKKEYSIIIVTHNQAAKISDKTAFFLNGEVEYDSDIIFSNPSDKRTEDIYITGR
FG"
complement(116778. .117662)
/gene="BH2992"
complement(116778. .117662)
/gene="BH2992"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter (permease)"
/protein_id="BAB06711.1"
/db_xref="GI:10175614"
/translation="MKYVDANVVQKMSRVLINQAKGLFELATIFGLVLAFLIAR
VLADGASWINLDFLCGRHSSDEKAGIMALLGTMLMLVTPVTMFLGVGTAIYLEC
YAKKGRLOQSETNINSLAGVPSIVFGILGTVFTMDLGSVWLAGLTMSLVLPI
VVASQEARAIAPSHLSEASFGLGATKQOTIRKNIILPAALPGILTGAILSLRAIGET
APLIVILGIPALLIPFGGLMDRFTLIPQIYYVITDSALVAEYANLAATAIILLVL
FLNLTATIIIRNKFQRY"
complement(117664. .118620)
/gene="BH2993"
complement(117664. .118620)
/gene="BH2993"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter (permease)"
/protein_id="BAB06712.1"
/db_xref="GI:10175615"
/translation="MNNSELNVRGAVDVREMIENRQTKSLANLAELKIPKCLLVIA
TVSILTTIGLITLLEHTEFFSRIPVDFGTGVLKPLSQNPFVGLPPLTGLIST
VAMVVAIPIGMTAIFLSEYASDLRRTLKPMLEVLAGIPTIVYGFPAFTVFLLR
HVVPGLATNIPSGIVMGIIMIPNVASLSEADSSVFNAMREGALAGATKLEVTK
VVIIPAISGIVASFVLGSRAGIETMIVTIIASGSSKNFTEDVTQSMQMTATYIVETG
GDAPAGSTIYSLYAVAMTLFVFTFLMILLARYISRRFREY"
complement(118840. .119847)
/gene="BH2994"

```

```

CDS
complement(118840. .119847)
/gene="BH2994"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter (phosphate-binding
protein)"
/protein_id="BAB06713.1"
/db_xref="GI:10175616"
/translation="MNVKFKFMLLTMTVLMVIVAACSGGEDTQAPGDETTAEPEET
TEATGSDLEGSVVVDSGTVYPLMAAIAEYMISEQPNVSEVSRAGTSAGFGKFL
VENGTDFNDASROJKDEKETAELGIDVQEFKVALOGLFVIHPDNDWATELKEEL
ISIFAADGGVEKSDIRPEPDEINPMGNENHGTYEFFEVIIMEODLVDVNVNQ
EYSLVNLVSEDPAIAFFGFGYVNNODOLGAVHVDGPGVPENLETAECDGYAP
FTRPVFTYLVNHNHAKPEVLDFALFYVNNVNFAGETGFAPLPEEVEQVYVNEGL
K"
complement(120124. .121395)
/gene="BH2995"
complement(120124. .121395)
/gene="BH2995"
/feature="BH2995"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06714.1"
/db_xref="GI:10175617"
/translation="MNSLPVAILGGPVGVLAAAHLAVRDVPFVLFEGKSEVGASILE
WKHVSLSFPMKYNIDRAAKSLLEKSEKPEAGDRPTGEQLVDRDYLOPLAETKALPY
ITYNPVQSIYRKIDIKDHNRENTPFIRYKODNDOKTLRARAVIDATGTFSPNP
IGAGVPAIGEEBAKDSIYVIPDVKGKDSRYEGKTLVVGSHSAIHTLLDLAQLK
EAPSTRILLWAIRNSPNGLFGGQEDALPARQOLGALAEALLSKQKVELISFEME
IDRVGRLSVGVQVNESPLSLVDIANTGARTVYDHRIRYQDLSKLESVPALA
PLDIPNHSCTVTRPHGKELRQPKNFYIAGVKSYGRAPTFLLATGYEQVRISVASL
AGSKKAERVELQLPKTVGKSDFLAKRSCC"
complement(121513. .121821)
/gene="BH2996"
complement(121513. .121821)
/gene="BH2996"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (ArsR family)"
/protein_id="BAB06715.1"
/db_xref="GI:10175618"
/translation="MDIELIPVPDEKIFEQKFKALADEKRLMLLVELCOHHEVCVCD
LQOMNMLPQSKLSYHLKILLDADLIERQKRGOMNYRLKDTVKALLSEQLCCVFRPQ"
122422. .122715
/gene="BH2997"
122422. .122715
/gene="BH2997"
/feature="BH2997"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06716.1"
/db_xref="GI:10175619"
/translation="MDVSEFVRYINLVHLSALLLVLDKTYEFIGMSAEWGRLLSQE
KRGYLOTWADDYIYALGANQKVRVGESIIISYHQAAIIVNSANDVAHVIRLK"
complement(122800. .123219)
/gene="arsC"
complement(122800. .123219)
/gene="arsC"
/EC_number="1.-.-.-"
/feature="BH2998"
/codon_start=1
/transl_table=11
/product="arsenate reductase"
/protein_id="BAB06717.1"
/db_xref="GI:10175620"
/translation="MNKKVIYFLCTGNSCRSQMARGWKYKLGDEWDVYSAGIEAHGV
NPNAVKMKIGIDISEQTSOTIDQELQKADLVVYTLGCHAADVCPATPSNKRHVHG
FDDPAKAGTDEKWAFFRRVRDEIGKRIKTAETGK"
complement(123247. .124557)
/gene="arsB"

```

CDS
complement(123247..124557)
/gene="arsB"
/note="BH2999"
/codon_start=1
/transl_table=11
/product="arsenic efflux pump"
/protein_id="BAB06718.1"
/db_xref="GI:10175621"
/transl_table=11
/protein_id="BAB06723.1"
/db_xref="GI:10175626"
/translation="MSSVIVASLIFFLTLVLVIMQPRGLSIGWSATIGAVLALLAGVV
DFADVQAVTGIVWNATLTFVAIIISILDEIGFFEMSAHMARVANGSGIKFLYIT
LLGAIVAALEANDGAALILPIVLAMVRAALKFKEAMVLPFVMASEFADTSLPLIVS
LVNIVSADFNFNIGFVEYATRMVVPNFALAAASIVWLILYLFKDKMPTKYDLSQKKPH
EAIKDKAMEFLAVLILIGVEYSEFVPIPVSVFAMPIAFYFYLIFRRASGKNKAL
NTTGIIGAPMAIVFISIGMYVYGLRNVGLTNILADIIEELAAAOGLFAGTAMGFI
AAILSLMNNMPTVMINALAIADNTMGIMREALIYANVIGSDLGPKITPTIGSLATLL
WLHLSKGMKIKTWGYFKVGIILTVPTLFTTLVGLYLWLMGIS"
complement(124586..124936)
/gene="arsR"
complement(124586..124936)
/gene="arsR"
/note="BH3000"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (ArsR family)"
/protein_id="BAB06719.1"
/db_xref="GI:10175622"
/translation="MASVKQOLEVATAVKVLKLGDKTRLTMTWALMLEDECCVCFVE
LFQMSQPSISQHLRKLRDVGLIEERGQWIFFRVNNKHEAYDMVASLLNHIPQTER
LRELEKGLRVCD"
complement(125094..125654)
/gene="BH3001"
complement(125094..125654)
/gene="BH3001"
/EC_number="2.3.1.79"
/codon_start=1
/transl_table=11
/product="maltose transacetylase (maltose
O-acetyltransferase)"
/protein_id="BAB06720.1"
/db_xref="GI:10175623"
/translation="MRTEKMKLAGERYKAWDPVLVDKREARLRLTFLNQTTETEEK
QRTELIKELFGSMGESVNIETPCFDGYNIHVGNNFANFDCVILDVCEVRIGANCM
LAPGVHITYATHPIHLPERVEPEYKGPVTIRNNVWIGGRAIVNPGVTIGNNAVIASG
SVTKDPVENVVAGNPAKVIQITDV"
complement(125736..126470)
/gene="BH3002"
complement(125736..126470)
/gene="BH3002"
/note="BH3002"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06721.1"
/db_xref="GI:10175624"
/translation="MKLTIVIGYHGYPCREGATSGYLLEQGETKLLDLCGSGVLSHVQ
AYCPLELDGIVLSHYHQDHMADIGVHYARLLQAGOGRIKKOAT IYGRDDDPETFS
LSYKHVVRSETYDGLNPKPIVGPQFSPFKHPVPCVAMRIEVEGOSLVVYADTSYA
ELANFAGCDLLIAECSAYGGTDLSDYCHMNSFDVGLAERTNARQLLLHLPHHGOH
EQLVAEVKTYQGEIFLANTGWKEW"
complement(126495..126671)
/gene="BH3003"
complement(126495..126671)
/gene="BH3003"
/note="BH3003"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06722.1"
/db_xref="GI:10175625"
/translation="MANFNQLTKEREAIQIODKINQTESEODRRKLIQTLLILEK
SRUKSENKKEQS"
126936..127253
/gene="BH3004"

gene
complement(130732..131880)
/gene="BH3008"
complement(130732..131880)
/gene="BH3008"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06727.1"
/db_xref="GI:10175630"
/translation="MKAIKNGTILTGGETIKQGTLLIQDGKVLLEVQQRVDTPANAKL
IDATGKYVTPGLIDVHPHIGVSEGLRGTGHDENETTAATPYIRADIGIYPDFQGF
DARQSGVTVQIMPGSANVIGGENVTVKTAGTVVDEIMIVRAPSGMKAAFGENKPNVHG

CDS
126936..127253
/gene="BH3004"
/note="BH3004"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06723.1"
/db_xref="GI:10175626"
/translation="MFDQDSIYIVGEAKSPSNPITQYKIFFIGFVVDRKTGKIVDA
DVSSILPTSTFIRHLLIGQNILETITQIEDRYHGSSQKAIIVALRNASQRFQOL
HTV"
127394..128560
/gene="BH3005"
127394..128560
/gene="BH3005"
/note="BH3005"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06724.1"
/db_xref="GI:10175627"
/translation="MIENGFLFLSILICMTALIVNAEYKIGGKFKYVPAIVLIVLGA
AFMNTFGLFGDSELSNWSYGVNRNALIPAMILLMLLQCDLRKIILGPKLLITFVAA
FSIVLGFLLYVMQSFDDTWFRAFALAGSWTGGSANMVALQDIIIAVPTETFGYAL
IMDTINYSFWVMWFWLVPFERFNRWTKADTSKLESMSQBIATAATVDEKREPTFVH
MIGLGLSFLTAALATVIGENLPOIGTINAMTWLILVIGVILLGALTPKIPSGM
DIGRVMLYTIVAITASGADFSSEIGPIVYITAGFMVLLFHLGLFGPAKLPKLDLFT
GVASLANIGMVSAPVLGAFNRALIPGVIMALIGGFMGTWFGVLTAEILSRL"
128575..129654
/gene="BH3006"
128575..129654
/gene="BH3006"
/codon_start=1
/transl_table=11
/product="muonate cycloisomerase"
/protein_id="BAB06725.1"
/db_xref="GI:10175628"
/translation="MNIRTFSIHQBTVALKEPFTALRTVTVEIVHVIITLESAGIG
RGSAAPTPTIGESKSMIAIINGPIMTALMARSMFOEALQTVONSCAGNTSAKAV
DIALHAYAQSLGISLIQLLGGASIPLENDMTVSLDEKIANEQMLRHLNHGFTQVKI
PVKNGVEDERILHLSHSAEKGVIFRIDANOAPKEAQQVIRQLEAQAVPVQFVEQ
KVAADLEGRLYRDHVTATPIMADESCTTTRTAALRIIOQAADMNLTKMKCGGLNEA
RYADLAQAADLPCMIGSMMESSWSVAAALASTHPNITMVDLDAPLWLMEEPSPH
LLYEKGTVFVHEPSPVKAPSSLNN"
129671..130681
/gene="BH3007"
129671..130681
/gene="BH3007"
/codon_start=1
/transl_table=11
/product="polysugar degrading enzyme (alpha-amylase)"
/protein_id="BAB06726.1"
/db_xref="GI:10175629"
/translation="MKIKWILMFLVVPVPSFGSTTYAKDPTKETTAYVDVSVLVV
WVEINARDVDPISLTNPDPKWKNTMYEKLWLGKLEQALYGMKMHILEEQD
WKTAVVQDPTPREAGYPGWPKQELIQSKFOOLRRRPATVYTAFTAFYISQOIT
KEENEVSFNPLVLAARDRVLVMTPSDGTKEFATDVEIYOSEDEIPATADLLE
TAEFLDPLYLWAGVSGFGDCSGFTHTITRAHGIIITIPRDSVQVARTGYVERREHLOK
GDILVFFARNGGTGAVHHVGMVIGDGKMIHSPNSASTVEITVIDESSYAGSYHSARRYL
EE"
complement(130732..131880)
/gene="BH3008"
complement(130732..131880)
/gene="BH3008"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06727.1"
/db_xref="GI:10175630"
/translation="MKAIKNGTILTGGETIKQGTLLIQDGKVLLEVQQRVDTPANAKL
IDATGKYVTPGLIDVHPHIGVSEGLRGTGHDENETTAATPYIRADIGIYPDFQGF
DARQSGVTVQIMPGSANVIGGENVTVKTAGTVVDEIMIVRAPSGMKAAFGENKPNVHG

KGKLPATRMGVAAALFRQOLYQAODYLKRPEKGELEKRDLCMEQLAKVLKRKEIDPLRVH
AHRDDTIATLVRWDEEDLEATIECHTEGHLIADHIAKYNKVSVGPTMTSRTKQELV
NKGWHTLIELEKHGIPFISITDHPVTINHLVTSIAQAVRAGLSEQSAFAQAITLRAAE
HIGVEQVGSLEKKGDDGVFVSDHPFTATAQVEATYINGKCVTGSHC"
complement(132008..132292)
/gene="rpmA"
/gene="rpmA"
/gene="rpmA"
/note="BH3009"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L27"
/protein_id="BAB06729.1"
/db_xref="GI:10175631"
/translation="MLKMNQFFASKKGVGSGTKNGRDSISKRLTKRADGQLVTGSGI
LVQRGRTRIPGVNVGKGGDTLFAKVDGVVKEFRVGRDRKQVSVYPKAQ"
complement(132298..132639)
/gene="BH3010"
/gene="BH3010"
/gene="BH3010"
/note="BH3010"
/note="BH3010"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06729.1"
/db_xref="GI:10175632"
/translation="MIDVFNERNQNDIVFTMSGHADAGPYGDLYCAGASAVAGLT
VNAILCALQVELVTEMENEGGFLRCRVPNDLEETTFEKVOLLLEGMNISLQSAESYG
EHIIQIERIRR"
complement(132648..132956)
/gene="rplU"
/gene="rplU"
/gene="rplU"
/note="BH3011"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L21"
/protein_id="BAB06730.1"
/db_xref="GI:10175633"
/translation="MYAIIETGGKQVKVEGQEIYIEKLDAAEGESVNFQKVVLVGGD
DVKVGSPYEGASVTAKVEKHGKGKIIVFKMAKKNYRRKQGRQPYTKVVIETKINA
"
133163..133333
/gene="BH3012"
133163..133333
/gene="BH3012"
/note="BH3012"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06731.1"
/db_xref="GI:10175634"
/translation="MKKELTMKIFLRYNNFAAHLPSVARSSFTPSQWKKRKSASL
PVKRGLLTLYLV"
complement(133299..133637)
/gene="BH3013"
complement(133299..133637)
/gene="BH3013"
/codon_start=1
/transl_table=11
/product="general stress protein"
/protein_id="BAB06732.1"
/db_xref="GI:10175635"
/translation="MSVKELTTVEQMOVEFETTTETPLVLFKHSQCQPVSAEAHQLH
ALLNGQLQDGVVLYKVIIESRPSVNSIAEDLGVKKHSPQAIIVVKDQEAIWKKSHWDI
KQDTLKAALV"
complement(133650..135152)
/gene="cafA"
complement(133650..135152)
/gene="cafA"
/note="BH3014"
/codon_start=1
/transl_table=11

/product="ribonuclease G (cytosolic axial filament
protein)"
/protein_id="BAB06733.1"
/db_xref="GI:10175636"
/translation="MKSILPMATRERRAAVAVENGQVIELMIERPEERIVGVNRYRGR
VVNVLPGMQACFVDIGREKNGFLYRDELLSPHLSNEEDEKKNISDFVOGEBELLV
QVTRGEFTGPRLTGVVAFPPYRVYVNPQAGYGVSKKIFPEERERWRSEWQOLLL
NDEGLIIRTACKKEVRHEAQEILFFROMWELQKRGKAKAPSLLYOSSGLIERILR
DFPLDEVGIEVDDREYHFIIKQLLASYPNMESRVFLYKGRNFSHLKVEKEIEKGL
RRVWLKNGGYLMDHAEALTVIDNSGKFICKQDLODTVVKNIEAKETAHOLRLR
DISGIIIDFIDMKREDDQQLVHTLEQALKTDRTKTNVIGMTGLGVEMTRKKVROS
LQASLKPCTCGTCGTQVLSDEAQAFKIERELMEYRGSDAEIIVLELPSHAPVLFGQ
DGHKLRETAIGSKILLAVPAKTLKEQYETIRYIGSLEEQFVERIKKR"
complement(135337..136206)
/gene="spoIVFB"
/gene="spoIVFB"
/note="BH3015"
/codon_start=1
/transl_table=11
/product="protease (processing of pro-sigma-K to active
sigma-K)"
/protein_id="BAB06734.1"
/db_xref="GI:10175637"
/translation="MISYLAFLKKMKVNPFLFWFVIGMVLGTGYREVLMVFVIVFVHE
MGHAVAHHFTWRIRKTELPLFGGVAEVEEGNRSFKEEVIIITAGPLQHVHMMAAAY
GLVALEMWNTSYQQFLWYNTVILAFNLPLPLDGLRMQLLCCMRFPWYRALRRTY
GLSFLFTVAVCAAILYPHLNLWLVTFLVNNYLEYKQCHYRFTFLTEREYIGL
SQQKPHLPVVVSAHQEVSEVAKQFRRGVIHTIIVNDAKQKRTHTLPETEVLRALFEA
KLPRAPLSTIIAK"
complement(136199..136981)
/gene="spoIVFA"
complement(136199..136981)
/gene="spoIVFA"
/note="BH3016"
/codon_start=1
/transl_table=11
/product="inhibitor of SpoIVFB"
/protein_id="BAB06735.1"
/db_xref="GI:10175638"
/translation="MARELOKLRKRLARRRINERTQKSRMVPLMFORHEEAAREEP
EVVYVPERNEPRASTDRGILLFTLISICLFLVIAVFOSSNAAPLOPLRAGVQAPAE
QEFOFAALAHWYEQFCQPLALIESNQDFALGDQDDQDERWVFAVPASGVITESFEQN
GRGLVETGVGAEEAAKSGYVIEINNSEKGMVTLLQHYDGSSEIYGMIDNVAVNVYD
HIEAGTKIGTASRAEDGEGKGVYYPALKOGENYVDPDSVIFRD"
137084..137680
/gene="moba"
137084..137680
/gene="moba"
/note="BH3017"
/codon_start=1
/transl_table=11
/product="molybdopterin-guanine dinucleotide biosynthesis"
/protein_id="BAB06736.1"
/db_xref="GI:10175639"
/translation="MNHLDRIGVILAGKSTRFRGRPRAMVQYQRYRYEASLDALQHH
TDQVLIISHPTLTNQFRNEGIRVIEDPRYQCGPLAGIFTATEKESAYWYLPAPCD
TPFLKIEYDILFSYLDQEPDKKAVIPVWNGKQPLIGLYHRSCLAPTQSLLSEQQRYK
VGFLFQLVDTLFEVDKAFEOHLHSSSFVNINRPEDLSEWT"
complement(137885..138121)
/gene="moaB"
complement(137885..138121)
/gene="moaB"
/note="BH3018"
/codon_start=1
/transl_table=11
/product="molybdopterin converting factor (subunit 1)"
/protein_id="BAB06737.1"
/db_xref="GI:10175640"
/translation="MIEILLFAELSEKAEFDRILLAGEKTVGELKMWAKEIFALGS
LEQIMVAINEEYANDHMTVKAGDVVALIPPVSGG"
complement(138118..138588)
/gene="moaE"

CDS
complement(138118. .138588)
/gene="moaB"
/note="BH3019"
/codon_start=1
/transl_table=11
/product="molybdopterin converting factor (subunit 2)"
/protein_id="BAB06738.1"
/db_xref="GI:10175641"
/translacion="MTSKRAIMEQPIVIOQVDOVVHPHAGAINFTIGTVRELTKCK
RTLYLOYEAVPSMAEKLAIOGPIORPWARVAITHRIGRLTIDVAVVIAVSTH
RADSYEASRAJERKEIIVPWKKEHEDGEMWGDQLETTSPGAPKMEEEE"
complement(138591. .139100)
/gene="mobB"
complement(138591. .139100)
/gene="mobB"
/note="BH3020"
/codon_start=1
/transl_table=11
/product="molybdopterin-guanine dinucleotide biosynthesis"
/protein_id="BAB06739.1"
/db_xref="GI:10175642"
/translacion="MKLIQVVGFNSGKTTLLEKLITQGRAGMKLATIKHHGHSPL
TALDKKDSYKHRQACGAVASVSSNGELQVOMSKPAPWRLEDPSLYRSLHLDGVFIE
GYKHLFAKIVLLKDEKVVLLRELQKVIIVVWDAALQIQIVDPVFTIHEERYL
SYLINWLEV"
complement(139097. .140368)
/gene="moaA"
complement(139097. .140368)
/gene="moaA"
/note="BH3021"
/codon_start=1
/transl_table=11
/product="molybdopterin biosynthesis"
/protein_id="BAB06740.1"
/db_xref="GI:10175643"
/translacion="MVERRVPIPVHEAVEQVAFAKERAMWVPLEADGRFTLAQPLV
ADHNPFKDSPLDGRFAIRAVDTDATRETPVRLKIIIEIGAGQVAKQVPSQEAIRI
MTQAQPEGDVAVVMEFLTREISEAGSEFTEVRSFQPGDNI SFCGDAKREGSPLVKE
GROIQIKALLATFGYKVPVYKMPKIGIFATGTTELLEVDPLVCKIRNSNAYMVA
SOYKIGAEVPIYGLKLVDDFDQCYDATALADVDVLTWGGVSGVDYDYLPAIYEKL
GAVNLEKIAMRGSVTVTAEREGQLLFGLSNPSACFVGFEYLLVRFPIHVGCSEOP
FLARKATLGRDPKPNPFTFRVRSQVEFFKIDRLIIVKTLGDKSGVVVTSUANGLLIM
LPQGTGRFQAGDEVVDVLLKGGGPMPLSEA"
complement(140402. .140914)
/gene="moaB"
complement(140402. .140914)
/gene="moaB"
/note="BH3022"
/codon_start=1
/transl_table=11
/product="molybdopterin precursor biosynthesis"
/protein_id="BAB06741.1"
/db_xref="GI:10175644"
/translacion="MSVEQHRKEAPKTIVRCMILTISTRTYDIDKSGALLRELEAKG
HQVTEYQVDEXYSQIOWHLKVAAGRADIDAILLNGGTGIALRDTTYEAVRDSLDKEM
PGGEYFRLLSFQFDIGTAAILSRAIAGVRDGAIFSMGPGSTGAVKIAMTRIIPELG
HVREILKDR"
141095. .141592
/gene="moaC"
141095. .141592
/note="BH3023"
/codon_start=1
/transl_table=11
/product="molybdenum cofactor biosynthesis protein C"
/protein_id="BAB06742.1"
/db_xref="GI:10175645"
/translacion="MRDFTHEQGRAKMVDISEKSTITERTAVAKSSVEVNEKIYEKI
IGKIRKGDVLVAQIAGVMAAKNTSOWIPMCHPLSLTGVDISFEWEETRNHYLLHTE
ASVKTGRTGVEMEALTAASATALTVYDMCKAIDKGMVIGPTIYLVKKTGGVHGDFERS
SSQKS"
141681. .142061
/gene="BH3024"

gene
complement(142134. .142649)
/gene="BH3025"
complement(142134. .142649)
/gene="BH3025"
/codon_start=1
/transl_table=11
/product="general stress protein"
/protein_id="BAB06744.1"
/db_xref="GI:10175647"
/translacion="MGKKIAVVTNLNLFEDSEYTEPVKAFKEAGHSIVTIEKEKGIKVK
KGQESITIDASIDEVTEDAFDALLIPGGSFPDILRADDRFVAFKAFADAKKTINA
ICHGPELLINAGVLKGRDVTGYKSIADVLNAGANFYDQVVVCGGNLVTSRPTDDLP
AFNRESUNVLT"
complement(142786. .144363)
/gene="BH3026"
complement(142786. .144363)
/gene="BH3026"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transpoter
(oligopeptide-binding protein)"
/protein_id="BAB06745.1"
/db_xref="GI:10175648"
/translacion="MKQWLLIWLISAFALVGCASEPSASQGGENEDARKGNALLIDL
MSYVTDHPDHPGANNESLYNMSTMYDVLVYDLSLEIQLPLAESLWMDTITWEAKUR
QPVTFDDSPFNAAEVKARLERVIDPDIGSPVAFLEDMITEVEVITYVHIITEYFP
APLPASHAHPGWHMISKEAIEADYEAMANGEPEPSYINHPVGTGFTYRGEAQNGEYV
KVKYNEVWREQAQSDALTFKVVPEDLTVRALETGEADIIYVNPNDVNRIRBOHNG
TKVLKSPSNLTLYLGENTTKEPNDPLVROALIAKIDQEGILINGILNGMALPAKGLA
TVFGYDQVQTSAAKKQDVEEAKALLDEAGYNGFETTIVSNDTRAFMDIAVFOQELQA
IGTVDQVETSTFLEAGNGDTEMFVSSWGTVDLDADYGLYPMFHSENAGFPNGRT
FFANEELDLLERARKETDEATRELYKDAQIIIVDEVPLVFLYHSELLAGLGEQVEG
FWQYPSIFYMRDAVKR"
complement(144800. .145594)
/gene="minD"
complement(144800. .145594)
/gene="minD"
/note="BH3027"
/codon_start=1
/transl_table=11
/product="septum site-determining protein"
/protein_id="BAB06746.1"
/db_xref="GI:10175649"
/translacion="MGEAIVITSGKGGVKTTSANIGTALASGKKVCLVDTDIGUR
NLDVWMLGNRIIYDLVVYVGRCLKQALIKDKRFECLNLLPAAQPKDKSAVTPEQM
KEVEELKQEDYVLIDCPAGIEQGFKNAGADKAIWTTPEISSVRDADRIGLLE
KEEVAEPRLVYVNRIRGHMKNGLMDVEIVSILAIELLGLIVVDENVIKFSNKGPEI
ALHPDSKASVAYRNIAARRILGETVPLNSFQEKGVLAKIKSFQVR"
complement(145617. .146315)
/gene="minC"
complement(145617. .146315)
/gene="minC"
/note="BH3028"
/codon_start=1
/transl_table=11
/product="septum site-determining protein"
/protein_id="BAB06747.1"
/db_xref="GI:10175650"
/translacion="MTTQKQAVTIKCTKQGLTFHLLDRCSFDSIVGCELAELKSSKH
QMEDQPRIQKVDVGYRILYVEQARRHQELITDGRNLDVEEFVSQWTKKEAEKKRKE
AQIVSAKVIRSGQVFDGDLILLIGDVPFGGTIRASGNIFVIGSLRGIAHGYKNG
EAVIATSHMAQPLRICEQIFHWSKEQQAQDRIMECAYIASDTQLDRVQKLKLI

CDS
141681. .142061
/gene="BH3024"
/note="BH3024"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06743.1"
/db_xref="GI:10175646"
/translacion="MQPFRVTLVEDEPSSHATLIQYHLNQLCAEVTVHPHSGSAFFOHR
QLSTDLIVSDOLVDLISFLDIDVKEQTKQPSVLITLTGRHELIESSEHNLSYLOK
PFAISELRAAIDYHKPSKGMTMV"
complement(142134. .142649)
/gene="BH3025"
complement(142134. .142649)
/gene="BH3025"
/codon_start=1
/transl_table=11
/product="general stress protein"
/protein_id="BAB06744.1"
/db_xref="GI:10175647"
/translacion="MGKKIAVVTNLNLFEDSEYTEPVKAFKEAGHSIVTIEKEKGIKVK
KGQESITIDASIDEVTEDAFDALLIPGGSFPDILRADDRFVAFKAFADAKKTINA
ICHGPELLINAGVLKGRDVTGYKSIADVLNAGANFYDQVVVCGGNLVTSRPTDDLP
AFNRESUNVLT"
complement(142786. .144363)
/gene="BH3026"
complement(142786. .144363)
/gene="BH3026"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transpoter
(oligopeptide-binding protein)"
/protein_id="BAB06745.1"
/db_xref="GI:10175648"
/translacion="MKQWLLIWLISAFALVGCASEPSASQGGENEDARKGNALLIDL
MSYVTDHPDHPGANNESLYNMSTMYDVLVYDLSLEIQLPLAESLWMDTITWEAKUR
QPVTFDDSPFNAAEVKARLERVIDPDIGSPVAFLEDMITEVEVITYVHIITEYFP
APLPASHAHPGWHMISKEAIEADYEAMANGEPEPSYINHPVGTGFTYRGEAQNGEYV
KVKYNEVWREQAQSDALTFKVVPEDLTVRALETGEADIIYVNPNDVNRIRBOHNG
TKVLKSPSNLTLYLGENTTKEPNDPLVROALIAKIDQEGILINGILNGMALPAKGLA
TVFGYDQVQTSAAKKQDVEEAKALLDEAGYNGFETTIVSNDTRAFMDIAVFOQELQA
IGTVDQVETSTFLEAGNGDTEMFVSSWGTVDLDADYGLYPMFHSENAGFPNGRT
FFANEELDLLERARKETDEATRELYKDAQIIIVDEVPLVFLYHSELLAGLGEQVEG
FWQYPSIFYMRDAVKR"
complement(144800. .145594)
/gene="minD"
complement(144800. .145594)
/gene="minD"
/note="BH3027"
/codon_start=1
/transl_table=11
/product="septum site-determining protein"
/protein_id="BAB06746.1"
/db_xref="GI:10175649"
/translacion="MGEAIVITSGKGGVKTTSANIGTALASGKKVCLVDTDIGUR
NLDVWMLGNRIIYDLVVYVGRCLKQALIKDKRFECLNLLPAAQPKDKSAVTPEQM
KEVEELKQEDYVLIDCPAGIEQGFKNAGADKAIWTTPEISSVRDADRIGLLE
KEEVAEPRLVYVNRIRGHMKNGLMDVEIVSILAIELLGLIVVDENVIKFSNKGPEI
ALHPDSKASVAYRNIAARRILGETVPLNSFQEKGVLAKIKSFQVR"
complement(145617. .146315)
/gene="minC"
complement(145617. .146315)
/gene="minC"
/note="BH3028"
/codon_start=1
/transl_table=11
/product="septum site-determining protein"
/protein_id="BAB06747.1"
/db_xref="GI:10175650"
/translacion="MTTQKQAVTIKCTKQGLTFHLLDRCSFDSIVGCELAELKSSKH
QMEDQPRIQKVDVGYRILYVEQARRHQELITDGRNLDVEEFVSQWTKKEAEKKRKE
AQIVSAKVIRSGQVFDGDLILLIGDVPFGGTIRASGNIFVIGSLRGIAHGYKNG
EAVIATSHMAQPLRICEQIFHWSKEQQAQDRIMECAYIASDTQLDRVQKLKLI

gene

RPNLATFDEMVEQ"

complement(146437..146961)

CDS

/gene="mred"

complement(146437..146961)

/gene="mred"

/note="BH3029"

/codon_start=1

/transl_table=11

/product="cell-shape determining protein"

/protein_id="BAB06748.1"

/db_xref="GI:10175651"

/translation="MSRFILPGLMFFIEVLEGLFQLVITERFASEIMIVPRELIVML

VFQIYIGRSGALYGLIFGLDFVYTDLIGVYMGFALLGYGFALTYKWLQYSNLT

LQVLAFLVLFYFYQYQLIIGLTEDMTMFQHRLLPTLALNGAFAILMVLPMRK

LQVYLMKAFREK"

complement(146958..147839)

/gene="mrec"

complement(146958..147839)

/gene="mrec"

/note="BH3030"

/codon_start=1

/transl_table=11

/product="cell-shape determining protein"

/protein_id="BAB06749.1"

/db_xref="GI:10175652"

/translation="MPSFFSNKKLIVLVSIIIVLVALIGYSLDRRLTWPFRFLHDS

VGWQVSMRPAHLVAGFFFEVSELRLHYENQTLKARLEYAQISVERNVLKEENES

LRETLDEESLSDYMRNAVYFRSPDRWNEAIGINIGSOHGVEDDMVAVSSNGYLIG

KIORVSEFSRVOLLTDTDRNRYRSARIDGDKAGVPVGTIEGDDETGLLVSKRMLEMD

AEIEPQGLVTSGLGTVPKGIIEVERVEPEYGTQNAVIRPYADLLSLSDYVIVI

ERUSTAMDELIIEGGL"

complement(147853..148893)

/gene="mreb"

complement(147853..148893)

/gene="mreb"

/note="BH3031"

/codon_start=1

/transl_table=11

/product="cell-shape determining protein"

/protein_id="BAB06750.1"

/db_xref="GI:10175653"

/translation="MFGFSKSDIGDGTANTLVYVKGKIVLRPSVAVKRTDTGTI

EAVGNDAKNGRTPGNIVATPMKDGVIADFTATMLKYFIRQAOKNRSIFTRKPA

VMVCVSGITAVKRAVEDATKAGAKYATLEEPFAAAGADLPVWEPTGSMVVDIG

GGTEVAIIISLGGVTSQISIRVAGDEMDAIIQVKKTYNLMIGERTAEALKLEIGSA

GSPGVNDMIRGRDLTLPLKTIETVTAEEISKALSQVOTIIEAVKNTLEQSPPELA

ADLMRGIVLTGGALLRDLKVLSDETNMPVIAEPEPLDCVAGTGRALENHLFRS

KAGITSADRKS"

complement(149030..149728)

/gene="BH3032"

complement(149030..149728)

/gene="BH3032"

/codon_start=1

/transl_table=11

/product="DNA repair protein"

/protein_id="BAB06751.1"

/db_xref="GI:10175654"

/translation="MSNWSMTLRDVPNRPRRRLLEHGAHALSNORIVAILMRTCTK

NESVQLAQHVLCHFGLRLLREATVEELTSIHGIGAKAIEFCAALELGRRIHTMQE

MDRTVIRTPEDVSFRFVMEERMLFSQENFVCLYLNTKNQVLHKQTVTFIGSLNASIVHPR

EVFKEALRRSAALSLCLHNHPSPGDPREDIEVTHRLAQVKGKLTIGIELDHHVILGR

TFVSLKEKHLFPFS"

complement(149836..150408)

/gene="maf"

complement(149836..150408)

/gene="maf"

/note="BH3033"

/codon_start=1

/transl_table=11

/product="septum formation"

/protein_id="BAB06752.1"

/db_xref="GI:10175655"

/translation="MKPLILASGSPRRKQLLEQNMNVPFTVCKSTIDETFDPTFPPEDEV

gene

VQQLARQAQDVAKKHEDSFILAADTIIVVFOGRILGKPATEQEARQMLSQLSDQSHEV

LTGVALLHQGVETVETTVRFWPLTDTETIETYLOTGEPDFKAGAYIQGLGAYLVK

CDS

complement(150599..151657)

/gene="spoilB"

complement(150599..151657)

/gene="spoilB"

/note="BH3034"

/codon_start=1

/transl_table=11

/product="stage II sporulation protein B (endospore

development)"

/protein_id="BAB06753.1"

/db_xref="GI:10175656"

/translation="MDONKAKVSRILNGRKROPFOELVENNPASAAKKNDRSRKRPRED

DKVIDFHKRYEDIRNREPFWDDGNRRKSPKPYKRKKKKPYGNERFALKVKKVPVI

LIMALVSALIVGTSLSGIFLITFTADELPVASGEVQETVPVAVSGDASFPPELAVVQ

GGAFSTSEAGEAVIASLOEKRAALVESDVSVMFGLAQDRAQAQVAQASELQD

DTYIKSPAFTSDSKRLNEETIQWLHDMESSYQQLVSLVSDGLTAGGARITKESVEGI

RKEVERLHQQKTAFOELDESTASHLQVINEHLIQAGNHLASYVSTQGNALWKRTQOE

LLDALVOQLFSRETSS"

complement(151777..152265)

/gene="comC"

complement(151777..152265)

/gene="comC"

/note="BH3035"

/codon_start=1

/transl_table=11

/product="late competence protein required for processing

and translocation"

/protein_id="BAB06754.1"

/db_xref="GI:10175657"

/translation="MLLVFGWNMELLVALVSLFSIFVSDLVYMIIPDRILFFAI

LFVLFRATLAPLNPWDAGANVGYILLFLVLIYPNGIGGQVKKLFTVLGVFLGK

LTLVAILMSLSCGAIGGVGMGLRIKRREIPPGPFILILGSLTAYLYGDALIQWYVR

WL"

complement(152663..154366)

/gene="BH3036"

complement(152663..154366)

/gene="BH3036"

/codon_start=1

/transl_table=11

/product="two-component sensor histidine kinase"

/protein_id="BAB06755.1"

/db_xref="GI:10175658"

/translation="MNQOKQWLGLVGLVLFVLSLVILPRGDLTATEKISIFLLI

MILTASPIIYRGLVLTLPQWVSLAVELKKGILLIVEIVMTQFPAITTYLSAGVYGSIL

RIPWNTFILTSLVSAAYFVLGGVEVGRGLAUTFOQLVPFIFAYAVTFLVYVHTLIH

LFRVYLKANSSELKGMREVAISSIMPLGLMLLYLSEAVGLVAVFVLGSAFAGIG

FLQWYHKQSYTNRMKDVTEYVELNACNTRKLIHLFAVQVEQFLEADTVLVEGKQ

QELNLLGSLDDAENSHFQSDLYSKNVETGSPYTMLRREDGLDRIRLPIYQSV

VSLPIROKEDTYGVITVFSKNMYHFKKQRIHMLLEILYVIAVAKNYDILLETTKQESV

RDLPTGLYNRYMEALKKTVNEADPFESILLDLDFHKKINDTYHQSGNDVLEAVA

DRMRKICPDNSVTSRYGGEFVILLPHHGEDAQVGILLQDDFSSSILIRISDLVVG

ETVDDIKVTASFGIAEGKDKHAMDILIRKADRAMYNGAKSGRNRVATYSITKQHSLS

L"

complement(154399..155691)

/gene="folC"

complement(154399..155691)

/gene="folC"

/EC_number="6.3.2.17"

/note="BH3037"

/codon_start=1

/transl_table=11

/product="folyl-polyglutamate synthetase"

/protein_id="BAB06756.1"

/db_xref="GI:10175659"

/translation="MQNERQAIEMHSFLPFKIPKQLQRIEMWMLKLGHPKEKELNVVH

IGGTNGKSTVSFMRCLTEAGNVGTFTPSYTEHFRERATYKGEPTSSSELVRSCLN

IIDVPELKRFDLGSPTFEFVITVIAFDFAVNPVDIVLLEVGIGGLDSTNIVTPIL

SIITSGDHMMNVLGGTLESIAIEKAGIIKPDIPAIIVHVQKQPEAFRVIQDKANKVKAP

LFAYIDFHSLSKSVTMDKQFFSTQSQSGQTLTSLLETVMKHHQHNASMAIQAIQEUR

SMKHLYELSDHIVIRGIRHNTTWIGRFEKISEAPUTFVDGAHNEEGMKALAEALLHFPT

gene
CDS

```
YKYRFVMAVTKEMKALLAPYVDLPASFTFTTFEFERAQAKAQLYEQAPVYNKSICA
DWKVAIRRESCKANEMWICSLGISYISKVRAFLLER"
complement(156315. .158957)
/genes="vals"
complement(156315. .158957)
/genes="vals"
/EC_number="6.1.1.9"
/notes="BH3038"
/codon_start=1
/transl_table=1
/product="valyl-tRNA synthetase"
/protein_id="BAB06757.1"
/db_xref="GI:10175660"
/translacion="MTQDQDEMTKYHPQDIESKWYEWIEGKFFEATSDETKKPYTI
VIPPNTVTKHLHGAMDTAQDITLSLRKMGQYDTLWLPQMCHAGIATQAKVEAKVR
EESKRYDLGREAFTVEKTEWKEEYADFIRQWAKLGLSLDYSRERTLDEGLSKAVR
EVPVKLYEKLILYRGEVLIINWDPVTKTALSDIEVYIKDVQGYFYHMKYPLADGDGAE
VATRPETMDTAVAPEDERTOHILGKMKVLPITGREIPVADDIYDMERFGSGAV
KITPAHPDNFEIGNRHNLPRILVMNEDGTWMEKAGQEGMDRPECRKQIVKDLQEOG
VLFKIEHVSVHSGHSERSGAVPEYLSYTFWVRMKPLAEQAIELQTTENKVNFPDRF
EKTYLRMENIRDCISROLWGHRIIPAWYHKETGEVYVGHGAEPQDIDNNKQDEVDLD
TWSSALWPSTMGWPDDEAPDYKRYSTDALVTGYDIIFFWVSRMIFQGLEFTGOPP
FKDVLHGLVDAEGOKMSKSLGNGVDPMEVDIKYGADALREFLATGSSPGNDLRFYW
EKVESTWNGKTNWASRFALMNDGMSYDEIDLTEKSIADQWILRLQETIETVTR
LIDAYEFGVGRHLYNFIWDLCDWYIEMAKLPLYGEDEVAKKTRSLVAYVLDQTMR
LHLPMPFIETEEIWHLPHEGESITVASVPVKSEETFEQAMGDMELLKDIIRSVRNT
RAELNVPMSKEIELHIQAKNEVDITQLERGKHVIEKFCNPSTLVMGTAIKPEKSMSN
VLSGVELYPLAGLLDLEERIEARLEKENLEKEVERVQKKLSNQGLAKAPEKIVIE
ERKKEADYLEKRAAVARIKELG"
159412. .159597
/genes="BH3039"
159412. .159597
/genes="BH3039"
/notes="BH3039
unknown"
/codon_start=1
/transl_table=1
/protein_id="BAB06758.1"
/db_xref="GI:10175661"
/translacion="MRRPYKKPPWYLKIRAIQAQLLPLLIIFQFRTLFFTSFDVI
LNFLLFYCCMLIRLF"
complement(159697. .160722)
/genes="BH3040"
complement(159697. .160722)
/genes="BH3040"
/notes="BH3040
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=1
/protein_id="BAB06759.1"
/db_xref="GI:10175662"
/translacion="MQQAQLPLEAILYQDLYPEAVEDHGENVKKLYTKHGLPALKKT
QMTFSQADWCVYCMRKLDRLGVRKVPVLPTEKGEYTLTDQDTSFYVMPWIESIPYNA
RESSEERLADQLGVILRLTVKTPQPFKEOLEASYNELDRDWSRLWTNRFADAQERK
TYMSPYELTLTHFSFLHEMAHTAKSELAWEYETCLEKEKRYSLVCHGRUSRAHAFFN
ERHPLFNFERAALDTPARDLASFFRHSFPHALWSEEGHRWLYRYEQHLPLEEEK
RLHLAYLSFPDPVITYGIEAYQQSQGPPSEVKVKRLEKRLNLNLRVERFSKSLLPPTA
INESGLS"
complement(160697. .161566)
/genes="BH3041"
complement(160697. .161566)
/genes="BH3041"
/notes="BH3041
unknown"
/codon_start=1
/transl_table=1
/protein_id="BAB06760.1"
/db_xref="GI:10175663"
/translacion="MALVNLHKENYSYQEVVOLSEHILQIOTYGHKRIEYWDFFLLBW
HVARNDCQOPPHVFDVRMIRTKDGRTAIPEYHOWITVHDHMEGAVPLQLEENREAIW
GAFTGNMLARGEQTPYSASTQERLIIAPSSVETGGQWAAALVREGKRRIQOSEQLTAWFS
KLEPPTLPASLSDIRGYSWNGLFFFCPRKEPQKSYAPLRTFLFLFWEEYGTDSLKLK
LDMDDDPHPLTDGGYGLLLAECLXPYELQGVGKDVYKWEWKARKLVRELSGWLDEK
```

gene
CDS

```
KKVSDYATSTTS"
complement(161579. .162829)
/genes="spovID"
complement(161579. .162829)
/genes="spovID"
/notes="BH3042
incomplete
required for assembly of the spore coat"
/codon_start=1
/transl_table=1
/product="stage VI sporulation protein D"
/protein_id="BAB06761.1"
/db_xref="GI:10175664"
/translacion="MTQEPSKLSFIEETVWLNKGOEVRKILAMSLEPDIRIEEKD
HVYIKGRLRLGVEYPTTEERATIDPTSAIOBOVSFRSVEETIDDEKGSIOHFPFPI
VATPLRIQKLDLIIYVOVDSFDYDLRPSRCIOLTADVISICMTSEPPVSKQAKOSER
AADPLPKPQLEPIPTSFTEAKKPFVNEAGAWNEGVSQRQTFKQPADNENKLEE
DVVVNSEQAHEGEPIDESKQAQAKVEASTASVOQOEENHQQZFTVSYANQDETAEG
VNEGNATPEAMMGVEVDHYDSVEEENPTKIQIGVQREEVAIDEESQEDIGEDT
DEPKGHEENALYLTKMTNGEERFSKLMKCIIOGENESLDMIAERYDISTSHLIRFN
LTEERVEGQILYIPVPSHQSEK"
complement(163099. .164388)
/genes="hemL"
complement(163099. .164388)
/genes="hemL"
/EC_number="5.4.3.8"
/notes="BH3043"
/codon_start=1
/transl_table=1
/product="glutamate-1-semialdehyde 2,1-aminotransferase"
/protein_id="BAB06762.1"
/db_xref="GI:10175665"
/translacion="MKTPEKSSAANRAKPLMPGGVNSPVRAFKSVNMDPVFMERGGK
ANIYDIDGNEVIDVLSWGLIILGHADDQVVEKLEKTEKTSFGAPSELETKLAEIV
IERVPSIEVYRVNMSGTEATMSALRGYTRGNKILKFECCYHGHGDSLLIKAGSGV
ATUGLPDQSPGVPETVAQNTLITVPYNLDESVRIAEQFGDDLAGVIVFPVAGNMGVPPP
EPGFEELRLRTEENGTLILFIDEVMTGFRVGYHCAQAGFITPDLTCLGKRVIGGLPV
GAYGKREIMEQIAPSGPIYQAGTLSGNPLAMTAYETLVOLTEANYEYFDRLGDLRA
EGISAVAKEDIPIHYTSRAGSMVGFFFTDKRVKNFADASSDLEFFAKYFKEMLHLGV
SLPPQFEGMFLSTPKHTEADIDTVDAAARQAFKRLKK"
complement(164402. .165388)
/genes="hemb"
complement(164402. .165388)
/genes="hemb"
/EC_number="4.2.1.24"
/notes="BH3044"
/codon_start=1
/transl_table=1
/product="delta-aminolevulinic acid dehydratase"
/protein_id="BAB06763.1"
/db_xref="GI:10175666"
/translacion="MGODLTFKRHRRRLQRTAAIRNMVRETAKLHIDLIYPIPLFVKEGEG
VROQVSPSNVYQVLSLDQDAEVDIVSLGIPAILFLGVPAEKDAVGSAAVHEHGIVQ
QAIQKVESYDPLIVADIATCLQFTDGHGCVIEEGKILNDPLDLARTAVSAQAKG
ADIIAPSNMDGFFVAAIRAGLDSAGFTDVPVMSYAVKAFASYGFFRDAHAHSSPVFG
KATQGMPPANRLEAREASDVEEGADFLIVKPALSYLDIIRVKNKTEGLTPVVAYNVS
GEYSWIKAAASLQGWIDKSTVLEKLIISMKRAGADLILTYFAKDVARWLNEEK"
complement(165391. .166152)
/genes="hemb"
complement(165391. .166152)
/genes="hemb"
/EC_number="4.2.1.75"
/notes="BH3045"
/codon_start=1
/transl_table=1
/product="uroporphyrinogen III cosynthase"
/protein_id="BAB06764.1"
/db_xref="GI:10175667"
/translacion="MIAPVLKKGKGLVITRAKEQASSLSRLTEEAGGVPIEVPLLSFTK
TMDVYEVSGASIOSYEWIVTFSANAVREFNTWPROELKLPSPISAVCSSPTQRELSRY
GISVDLMPNEVVAESLIVETWVKRLKPCRVLPKGNLARPTVKVKGLETNGFIYDENTYI
ETKRLPBAKVMCLCAQKHRIEIIITPSPSAVDHFVTLDDREFLLHLLHLLIAYIGPIA
AARAKQRGPLVHIIVAKTYTGEGLVNAICDFYKEDG"
```

gene
CDS

```
gene  
CDS
```

gene
CDS

```
gene  
CDS
```

```
gene
complement(166149. .167084)
/gene="hemic"
CDS
complement(166149. .167084)
/gene="hemic"
/EC_number="4.3.1.8"
/note="BH3046"
/codon_start=1
/transl_table=11
/product="porphobilinogen deaminase"
/protein_id="BAB06765.1"
/db_xref="GI:10175668"
/translacion="MRKIIIGSRSSQLALQTQEWISKLKEAGAPDFEIKKIITKGD
QILDVLSKVGGLFVKEKEQMANKEIDIIVHSMKDVPSVLQEGFALGATTIRVDP
RDLISNEGTVLAEPLPAGSIVGTSSLSRQAQILAKRPDLKVRKIRGNIETRLKKEE
FDALILAAAGLRLMGWSNEIVTEYLDPLCVAPGQCALGIECRADDSDVLEMLKTI
NDDTAKTVLAERTFLHMEGGQVPIAGYATLTDEGEIELTALVGPDPGEQIFKEIL
RGKDPVLTGTMAQALMDQAKVLEQVKGLESQ"
complement(167120. .167935)
/gene="hemx"
CDS
complement(167120. .167935)
/gene="hemx"
/note="BH3047"
/codon_start=1
/transl_table=11
/product="negative effector of the concentration of Hema"
/protein_id="BAB06766.1"
/db_xref="GI:10175669"
/translacion="MNNIYPLTIIISISLFAFYFIDFLOHNRKVNRLAFWLLSVWLL
QTVFFIIRADLNRPLTIPTEGGLFFVAVVWVTLSLINRFRVDELVFFVNIIGFVM
MSFLPTPDGDPESLASLILISLIIHVSMILLSYAGLAFASFALYVLEHOMLKR
KMGKRLRLGSLTKLDQYAFHSTMAFPFLGLIIGFVWASIQFGELEPWFDAKVIS
SMVMCGAYVLYORVYKMRGYSMAALLNACFLILILINFLPGLSTFHLWD"
complement(167965. .169344)
/gene="hema"
CDS
complement(167965. .169344)
/gene="hema"
/EC_number="1.2.1.-"
/note="BH3048"
/codon_start=1
/transl_table=11
/product="glutamyl-L-cRNA reductase"
/protein_id="BAB06767.1"
/db_xref="GI:10175670"
/translacion="MHILMIGLWYKAPVIREIKFSQSELFOALHQLRQMSKILEC
TVIVSTNRELTYVADOLHTGRFTFTFLADWFKLPKDEFTPELTIRENDHAIEHLFR
VVTGLSDMILQILQVRNSETFIAQEQVGTGIFNHLFKQAITLAKRAVSETDQO
NAVSVSAVLELGGKIIFDDEKGVLLIGAGKMGELTAKHLHNGAEQVTVIINRTREK
AALAKRFLGVDRYNELTEAIVEADILISSTGATGVVVTSDMVSHALKRKRPLPM
VDIAPRDLDPALASHDDVLYLIDDLONIVQTNLBERTEAEKIELLIEELVEFRQ
WNLTLGVPIITALTALTKALTVQGETMESIERKLPNLTEREKKVIRKHTKSIVNOLLRD
PITRIKELANAPEREEALDLFTKIFALEELABQEKQKVKQAEQEWLAKKRPIITCME
KQSHVMVKS"
complement(169673. .170257)
/gene="BH3049"
CDS
complement(169673. .170257)
/gene="BH3049"
/codon_start=1
/transl_table=11
/product="GTP-binding protein"
/protein_id="BAB06768.1"
/db_xref="GI:10175671"
/translacion="MKVTRAEALDCVAKPEQYNSPLPEVALAGRSNVGKSSFINMKI
NRKRLARTSORPGKTQLNFYNTINDLHFVDVPGYGFYKAVSKKRDRAWGRMIETLQO
RGQLKAVLQIDIRHDPTRDQQLMYDLWLKHFEIPVITIIATKCKIPKQWPKHVKNIA
SVLEKDGDPVIVFSSETGTGKDEAWKAILHAIS"
complement(170254. .172578)
/gene="lona"
CDS
complement(170254. .172578)
/gene="lona"
/EC_number="3.4.21.53"
/note="BH3050"
/codon_start=1
/transl_table=11
```

```

gene
complement(172711. .174381)
CDS
complement(172711. .174381)
/gene="lonB"
/gene="lonB"
/EC_number="3.4.21.-"
/note="BH3051"
/codon_start=1
/transl_table=11
/product="ATP-dependent proteinase La"
/protein_id="BAB06770.1"
/db_xref="GI:10175673"
/translacion="MNMWTGILFVVQLFVVGIVGLYFWNLKLTORSQSVSDIRSRKEM
DQLRQLQSLSEPLSEKVPESFEDIVQKGEIRALKAALCGPNQPHVITYPGPVG
TKAAALQLQEAKNPSPFAHDVFDLDTTARFEDGIADPLIGSVHDPYIQGAG
AMQAGQIPQKQAVTKAHGMLFIDIEIGLHSIQINKLVLDRKLVLESAYSE
NQOIPIRHIHFORGLPADFLVAGTRMPEEIPPAIRSRCLVYFSLAREEIVTIA
KAVNRLGFKAEQVYEKISAYATNGRETYNIVQIAGMATTEGRHRYTVADVWVH
SSQLSPERQIPSEPKVGLNGLVPGNPGALVEVDVIPSAGSVYITIGIAEE
EQTGNQARSIRKSMAGSVENLVTLKMDPLDIDYDHIHFPAGPVPDGPAGVAI
ATGIYSAIKKEEVNDVTAMTGEISIHGTVPKGVGIKAVKAASGAKTFLIPKNEQ
AIQIEIEGITIIPMKHLDDVLSYALIREETEETETIPANQSFVSAPAPV"
complement(174826. .176100)
/gene="clpx"
CDS
complement(174826. .176100)
/gene="clpx"
/EC_number="3.4.21.-"
/note="BH3052"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease ATP-binding subunit
(class III heat-shock protein)"
/protein_id="BAB06771.1"
/db_xref="GI:10175674"
/translacion="MFKNDKGLKCSFGCKTQDOVRKLVAGPVVYICDECIELCTE
IVEEELGTEEEVEFQVPKPHEIREILDYVIGQDLAKLSLVAVYNHYKRNMSKS
EEVELSKSNICMIGPTGSGKTLQAQLRLINVPFALADTSLEAGYVGEDVENIIL
KLIQAADYDVERKAEKGIYIDEIDKVRKSNPSTIRDSVGEVQALKLLEGTTAS
VPPQGRKHPHQEIQIDITNIFICGGAFDGEIQIKRRLGKVVFGFSEVQKDELK
PGEYLSKVLPEDLREFLPEFIFGRPLPISLQDLDEALIELTKPNALVKQYKKL
LELDVNLFTDEALREIARKATERKTGARGLSIIEGLTLDVMFDLPARDVETCII
HDKCVTEGEPPIILKMDGSEVAMTKPEKSA"
complement(176627. .177922)
/gene="tig"
CDS
complement(176627. .177922)
/gene="tig"
/note="BH3053"
/codon_start=1
/transl_table=11
/product="trigger factor (prolyl isomerase)"
/protein_id="BAB06772.1"
/db_xref="GI:10175675"
/translacion="MTAKWKEKLENGEVLTVDVESAKVDEALDKAFKVVVKVNPQGF
RKGVPRKIFEROFGEVYALQDALDILLPAYAAAIDETGIEPVDPEIDIEOMEQGN
NLIFKATVYVKPEVQLGDYKGLEFEKEDTVSDVEDQELKSLQERQAEALVYVEERAI
QEGDTAVLDQFEGFVDGEAFEGGKAENYSLEIGSQGFIPGFEDQLVLKAGEKNDVEVT
```


gene
CDS
FPEEYHAEELAGPATFKVKIHDVKKKELPELDEFAKDVDEEVEESLDELKKLREKL
EKDRAHEADHEKRDTLIQASENATIDIEAMINTELDRTMQEFORLQMGNNLEMY
FOFSGOTQEOREOMKEDAEKRVRVNLTLEAIANOENLEASDEDEVEKELEKMAEYQR
SVDEIKSIFATOGTGDGIKADLKIQKAVDFLVEHSKAVS"
complement(178026. .179087)
/gene="BH3054"
complement(178026. .179087)
/gene="BH3054"
/note="BH3054"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06773.1"
/db_xref="GI:10175676"
/translation="MERYNTMSGKKTSSDNIVLYPGLVKRLTEKGMEALKKEDAKKA
LTFEQAQOYDAEPSOILFGKALCYVELGLKDDAVLLTQDMLRNDIGNYDIILQIHS
LLVOLGRYDEVMLDVLAEDELPAKHAESFYQLLHFSQMVEDSPTYIKELSEL
DLKETDIDNERSLEQWHFLQOMRKHNFPADIRTLLSFVSESQEPALRSALHMT
MHKVERKVRIRKPDQAEVLPVELNADQOFGNAVVEQLEARLAVENPSLFKMAQOL
WMOFLYALPFTPKPEQKDLWAGAVHAVHESNGLPFDDEEIIAGMYNCRVQQMLERGE
EIHVETETFGISGHSFR"
complement(179234. .179818)
/gene="leuD"
complement(179234. .179818)
/gene="leuD"
/EC_number="4.2.1.33"
/note="BH3055"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate dehydratase small subunit"
/protein_id="BAB06774.1"
/db_xref="GI:10175677"
/translation="MEPLKTHGTIAVLDRVNVDTQIIPKQFLKRVETGFGQFIHFF
DWLEFNGENDPDELNPHAKOGTILVAGHNFCCGSSREHAQWAKDYGRFVVIAPS
FADIFYNCFKNGILPVKLKQEVLTALMEKGKDTAYSLVNLQAIQVTSDDGEATFE
IDGYKEMLINGDEIGLTRIYEQIKQFEASRV"
complement(179843. .181261)
/gene="leuC"
complement(179843. .181261)
/gene="leuC"
/note="BH3056"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate dehydratase large subunit"
/protein_id="BAB06775.1"
/db_xref="GI:10175678"
/translation="MAPKTIIEKIWAHTVIGEGKPSLLYIDLHMVHETSQPAPEG
LRLAGPVRPDLTFATMDHNVDVDFRNIDQDIARKQIETLEANCKEFGIEIAGLDS
PNNGIVHVGPELGLTPGKTIIVCGDSHTTHGAFALAFGIGTSEVHVLATQTLWQ
SKPKTEVRYTGELAPSVSAKDIILAVIAKYGVDFGTGHVIEFTGEAIRLSMEERM
ICNMSTEAGAKAGLISDPSVTFEYLRGRPNAPKGEAFDVAIQOWEALATAGAVYDRV
LTHNASEIEPMVTWGTNPAOGTGVSOVYVPSDDAKDENERRAIKQSIAYMGLPEGPI
TEIAIGHVFISGCTNSDLRTAAELIKGRKVDAGVRALVPVGSQVQRKAAEKEGLD
EIFKEAGFEWSDGSCNCLGNPDTPVEGERCASTSNRNPEGRGKGARTHLVSPQMA
AAAAIAGHFYDVRTFQSEPQT"
complement(181289. .182380)
/gene="leuB"
complement(181289. .182380)
/gene="leuB"
/EC_number="1.1.1.85"
/note="BH3057"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate dehydrogenase"
/protein_id="BAB06776.1"
/db_xref="GI:10175679"
/translation="MEKQIADLPDGDIGTPEVTDAAIEVLOAVDRFGHTTFSYKALLG
GCAIDVSTPLPETLDVCRHADGILLGAVGGPKWDTLPGLRPEKGLGILGRKGLNF
ANLRPTVYSLADASTLKNDVIDGVLLIVRELTTGGLYGEPRRERGEETEVRDT
LLYLVTRGEMRIIRKAFELAVRNKHVTSDKANVLESSMRWEVYANVAEFPVETLLE
HMLVDNAAQRIIRKAPQFDVLVTENLFDGILSDASNVTSGLMPLPSASITSDGPLY
EPIHGSAPDIAGKGVANPLATIASCAMMLKYVSGLHEEAKTIEDIAIEAVLKQGYRTAD
IAKPGESSSTKAITDAVVEAIIQASVDIS"

gene
CDS
complement(182385. .183932)
/gene="leuA"
complement(182385. .183932)
/gene="leuA"
/EC_number="4.1.3.12"
/note="BH3058"
/codon_start=1
/transl_table=11
/product="2-isopropylmalate synthase"
/protein_id="BAB06777.1"
/db_xref="GI:10175680"
/translation="MKRISVFTDTRLRDGEQAGVNLNFEKMEIAKQLERGLGVDTIEA
GFPASSQGDFFQSVKAIATVTKSSVTGLARSYQKQIDAWEALKSAEERPHLFIATS
PIHMEKLRMTPEQVIEKAVESKYARSKFQWQSAEDASDFPLAQIIEAVTQA
GATVILNLPOTVGYTTPQEIKRQFYMKQNVPSIDKVSLSLTHHNDLCLMAVANSIAAQI
GDQADVCTINGIGERAGNASLEIAVALNIRKDHVETETGLILKEIKRTSSLVSKLT
GMVVPANKVWGANAAHESGIIHQDGVLNKQVTEIITPEMVGVSNSMWLGRKSGRH
AFTKKVQELGFTGTDEQLNNIFKAFDLADKKKEITEDDLFALMTBOTVCGEINHVEL
QTLOVNYGSLNTATITMKLPSGEVAEEAATGSGVEAIYNTLERLLDAPVKLLDTR
IQSITGGRDALADVYQMDYQGVSSVSGRTAHDVLEASAKAYLNAVNRITNRRKKYAEV
YGSKVEV"
complement(183919. .184941)
/gene="ilvC"
complement(183919. .184941)
/gene="ilvC"
/EC_number="1.1.1.86"
/note="BH3059"
/codon_start=1
/transl_table=11
/product="ketol-acid reductoisomerase"
/protein_id="BAB06778.1"
/db_xref="GI:10175681"
/translation="MAKVFYNGDINEGVLOGKTVAIIGYSGQGHAAQNLRDSGOEVI
VGLRPGKSWDKAAEDGFQVYSVREAAASRADVIMLLPDEHQTVYKNEIPELSEGT
LAFAHGNVFNQIYVPPATVDVFLAAPKPGCHLVRRITYVDGAGVGLVAVYQDATGQA
KDIALAYKMGNSARAGVIETTFQEEETDPLFGEQAVLCGCTSAVKAGFETLVEAGY
QPEVAYFECHELKLIYDMLVEGGLYMYISIDTAQWGDFFQAGPRVVTATKOAMKD
ILSDIQTGKFAKGWILENQANRPEFTAINEREKNHPLEVYVRELMMPFVAKSKGV
VGSAKN"
complement(184979. .185500)
/gene="ilvN"
complement(184979. .185500)
/gene="ilvN"
/EC_number="4.1.3.18"
/note="BH3060"
/codon_start=1
/transl_table=11
/product="acetolactate synthase small subunit"
/protein_id="BAB06779.1"
/db_xref="GI:10175682"
/translation="MKRTITVLNNTSGVLNRITGLFARRHFNIESITVGTENTPTS
RMTFVTVLDNOKNIEOVIKOLHKQVDVLKYKIDTDEAIVARELALIKVVASPOORGFI
AAVNPFRASIIDIGRESMTVQITGDHQKVETFDILLRPIGKELARTGITAKRGSK
KYDVEGNHITLIN"
complement(186420. .187226)
/gene="BH3061"
complement(186420. .187226)
/gene="BH3061"
/note="BH3061"
incomplete"
/codon_start=1
/transl_table=11
/product="acetolactate synthase large subunit"
/protein_id="BAB06780.1"
/db_xref="GI:10175683"
/translation="MSTNFKLKKREQKPEPQPMSSGSMLIQALAREGVDVIFPGPG
GAILPTDSELYKTGIRHILAREHQGAIHAAEGYARVSGKPGCVVTSQPGATNVVTGI
ADAMISDLVTVITGOVATTLIGTDAFOEADMIGITMPTITKHSYQVPTIEDLPRIAKE
AFHATITGRGPVLIDLPKDISLTGLENVDKVEHLPGCYOTVMPNRLQIRKLVEAS
QAKRPVILAGAVLGHAAQGLNMLNKNKPFQSSIHCUDSEAFRRIMNYI"
complement(187593. .189260)
/gene="ilvD"
complement(187593. .189260)

```

/gene="ilvD"
/EC_number="4.2.1.9"
/note="BH3062"
/codon_start=1
/transl_table=11
/product="dihydroxy-acid dehydratase"
/protein_id="BAB06781.1"
/db_xref="GI:10175684"
/translation="MRSDMIKKGIDRAPHRSLRAAGVKEEMDKPFIIGVCNSYIDII
PGHMHINKAEVAKEAIREAGGIPPEFNTIGVDDGIAMGHIGMRYSLPSPREIICDAE
TVINAHFQGVFFIPNCDKITPGMLMASVRTNVPSFVSGGPMCAGRTKDGKLSLAS
VFEGVAFSGSGMTREELLEIEQTACPTCGSCSGMTANSMNSLMGLMGLPNGCTI
VATSEARHOLIKDAAKHLANLLEKDIRPOLITKETIDDAFALDMAMGSTNTVLHTL
ATANEAEIYDLNRINEVAERVPYLCKISPASYSDMDVHHAGGVAAILKEICEIDIG
IHPDRTITGKSIYENKDAEITDDVVIRKKNPSPVGGLSLFLGNLAPNGAVIKVG
ADVPSIQIEGAIIVYNQSEAAQOGLNNGDVREGHVVIYIEGPKGPGMPMLAPTS
AIIIGLGTKVALIYDGRFSAGSGISIGHISPEAAEGGPIAFIENGDKIRIDLPRNT
IEWLVSDEIAKROEGWTEPEPKVKKGYLARYSKLVSANTGGVMKI"
189701..190051
/gene="BH3063"
189701..190051
/gene="BH3063"
/note="BH3063
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06782.1"
/db_xref="GI:10175685"
/translation="MLVENGRMKCKTEESDFYELLIOENLYPTLNHYHVGSPVHFAL
VPFKALIERHSHMNEKRLMKRLKRRKWDVLFESRRLATERGRLLVNDIVHILSQK
PSFSLFETKKGLSH"
complement(190642..191289)
/gene="BH3064"
complement(190642..191289)
/gene="BH3064"
/note="BH3064
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06783.1"
/db_xref="GI:10175686"
/translation="MSNETKPKPKFYKWWMLIAVIVIFAAVSGGDDSSSETAOEPE
PEAEVSDQSENESEPEEAEEDTEESAEEEDPIAGLGEALKVGDVVFTANGT
STAGSVGLVLAEAAGKFTLIVDVTIKNGSDSITVDSFFKLVGDVEYDSSSAGLY
ANEGADFTLTKLNPGLPLGKVPDYDQVLDSDILLNVQTFGTQOGEIKIQ"
complement(191478..191690)
/gene="BH3065"
complement(191478..191690)
/gene="BH3065"
/note="BH3065
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06784.1"
/db_xref="GI:10175687"
/translation="MPTGKALNVRCVLVEDTELEPKAIVETSKEIATITLSRVPIK
KGKNGLSLSPKCGIKGYMIIGSK"
complement(191840..191916)
/gene="trnQ-Arg"
/product="transfer RNA-Arg"
complement(191840..191916)
/gene="trnQ-Arg"
complement(191921..191994)
/gene="trnQ-Gly"
/product="transfer RNA-Gly"
complement(191921..191994)
/gene="trnQ-Gly"
complement(192322..192831)
/gene="BH3066"
complement(192322..192831)
/gene="BH3066"
/note="BH3066
unknown conserved protein"

```

```

/codon_start=1
/transl_table=11
/protein_id="BAB06785.1"
/db_xref="GI:10175688"
/translation="MKLLILSDSHGWSDELKAVADKHKRQEVDAIHCDCSELPRDDRA
LEGNNIVRGNCDFGDFPEFIKTVGDFNVTVTHGLYNVKMSXVSLTYRAEEVGAQL
VCFGSHVATSFQFNGIVFVNPGLRPNRKEQTYCLAYVRDDQIETFLDRDGHV
TDLQRTYLR"
complement(192859..193443)
/gene="BH3067"
complement(192859..193443)
/gene="BH3067"
/note="BH3067
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06786.1"
/db_xref="GI:10175689"
/translation="MKEWIVATKNKGVAEFAILKRGFSVKSLLDYPAIEDIEETG
STFENATIKAEIAERFQVPLADDSGLIIDALDGRPGIFSARYAGEEKDDQKNIEK
VRELQDIPWKARTARFHCSTIALARPOAETIVFEGTCEGYITTEPKGTGGFGYDPFY
VPSHDKTMAELTOEKNKLSHRVHALKOLDKWD"
complement(193440..194213)
/gene="rph"
complement(193440..194213)
/gene="rph"
/EC_number="2.7.7.56"
/note="BH3068"
/codon_start=1
/transl_table=11
/product="ribonuclease PH"
/protein_id="BAB06787.1"
/db_xref="GI:10175690"
/translation="MRMDERKADELRPISTEINVLHHPEGSVLISIGNTKVICTASLE
DRVPPFLRGQKGWITAEXSMLPRATEORNNRESTGKVSRTMEIQRLLGRALSII
DLDKIGERTWIDCDVLOADGGTFTTITGAFVALSMEKALANGAIQWPIIDDFLA
AVSVDGPKLGAILLDCYAEADAQAEVDMNVMTGKEIYELQGTGGEATFSHQLLTI
LALADKGIKQIIEQRNALGAIANRVDEAKQKKKEQSEA"
complement(194250..194378)
/gene="BH3069"
complement(194250..194378)
/gene="BH3069"
/note="BH3069
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06788.1"
/db_xref="GI:10175691"
/translation="MKRYLFLHDWKGYRPFSEVRKTIITFASKGSMNELSFSFYNT"
complement(194440..195537)
/gene="gerM"
complement(194440..195537)
/gene="gerM"
/note="BH3070"
/codon_start=1
/transl_table=11
/product="germination (cortex hydrolysis) and sporulation
(stage II, multiple polar septa)"
/protein_id="BAB06789.1"
/db_xref="GI:10175692"
/translation="MRKILRKGAFFVIIAASLAFACSGANEALNEMDTPPVNVE
EGESLEEEGVPEDEEREKEEGDEESAETVNRRELYLDDGVMPVQTLPLPK
SDGLVQLSQLEYLVEGGPVTLNPGFAVLPDPTSMVNLGDAVGVSKFETDVG
EKEQQLIASITWTLTQFENVRKVLQVNGHLEKMPVNGTPIGDGYSRANGIESGD
VVDVYNSAVTLTFLAQSGDQTYVYVQVTRRVNKNDSFATAVEELLGPMVTSPLVTD
FRNGVELLDEPKYENCYVTLNENEALLSOMQATAVSDEIINMLALTLETQDQGVKVAI
QYNGETNILEQGEFVAEPVSRPSQVNTGF"
complement(195714..196316)
/gene="BH3071"
complement(195714..196316)
/gene="BH3071"
/codon_start=1
/transl_table=11

```

```
/product="glutamate racemase"
/db_xref="GI:10175693"
/translation="MLVIACNTAAAVLDEVKELPIPVGVVQPGAISALKVTKNEH
VAVIGTGTIOSGAYEHTLKKINNRQVESLACPPFVLVERGIFSGPEALFTCSETL
KPLQKGFTLILGCTHYPLLPKPVIOEVMGPDQIISSGDETAREVSGLLYHKQRLRT
CNPAPHHPTTGDASKFRKLADVWLEMLNVPVGVETITLES"
/gene="BH3072"
complement(196525. .196980)
/gene="BH3072"
complement(196525. .196980)
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/db_xref="GI:10175694"
/translation="MERROVEQIEKSLRMVAETIVKOKGREILTHPELTPQFVALQWL
NEHGDWTIGELSKMYLACSTTDLIDRMEKNHLVERVKDQNDRRVRIRHLLERGKTI
IEVIRKREYIFDIIEKFSDEEVSMPFNLYLVYDEMOAKKEEKKRN"
/gene="ptsI"
complement(197102. .198820)
/gene="ptsI"
complement(197102. .198820)
/EC_number="2.7.3.9"
/gene="BH3073"
/codon_start=1
/transl_table=11
/product="pts system, enzyme I"
/db_xref="GI:10175695"
/translation="MHKKITGIASSGVAIAKAFIHLEPNVDIEKTSITDVSIEIKL
TAALKSKELRAIKQTEASMGADKAEIFAALLVDDPELVDGKIGKIENQMAAE
YALKVSDMFVSFMSQMDNEYMKRAADIRDVSKRVLGHIGVETGSLATIAEETVII
AEDLTSDAQLNKQFVKGFATDIGRTSHSATMSRSLIIPAVVGKVEYKTEKIQHGD
VIVDGEGIVNPTTEEVKAEKRAEKQKQWAKLVGEPTSKDGAHVLAANI
GTPKQDVGLANGEGIGLYRTEFLYMDRDLQDTEEEQFEAYKEVVORMDGPVVIRT
LDIGGKEUSYLQPKELAPFLGFRALRICLEKQDIFRQLRALRASTYGNLKMVPF
MATLEELQKAIQMEERDKLLSEGVDSYSEIYGMVEIPSTAVAANLFAKEVDFF
SITGNDLIIQVTAADRNERVSYLYQYPHAILRLVDMVIKAAHSGKVVGMCGEMAG
DEVAIPILLGLGLDFESMSATSILPARSOLLKLSKEELPFAQKALMLDTAEEVEQLV
KKTYLK"
complement(198823. .199086)
/gene="ptsH"
complement(198823. .199086)
/gene="ptsH"
/notes="BH3074"
/codon_start=1
/transl_table=11
/product="PTS system, histidine-containing phosphocarrier
protein (Hpr protein)"
/protein_id="BAB06793.1"
/db_xref="GI:10175696"
/translation="MAEKTFTITAETGIIHARPATQLVNKAGQYSSSEITLEYKQKAVNL
KSTMGVMSLGVGKAQVTTINAGSDEAKLGIEEVIKEGLGE"
/gene="gerE"
199712. .199936
/gene="gerE"
199712. .199936
/notes="BH3075"
/codon_start=1
/transl_table=11
/product="transcriptional regulator of spore coat protein
(spore germination protein)"
/protein_id="BAB06794.1"
/db_xref="GI:10175697"
/translation="MKGAEGFKPPLTKRREVFELLVQDKTKETAELQFISEKTVR
NHISNTMQXLGVKGRSQAVIELIRLGELEI"
complement(200214. .200690)
/gene="paiA"
complement(200214. .200690)
/gene="paiA"
/notes="BH3076"
/codon_start=1
```

```
/transl_table=11
/product="transcriptional repressor of sporulation,
separation and degradation"
/protein_id="BAB06795.1"
/db_xref="GI:10175698"
/translation="MLQEISIVTFNDTFFKDQNSPENWNTYLEKAFNLMQLEKELANPA
SQFFFYIFKNEVAGYLKINTDDAQSEELGEESLEIERIYIRKEFQKRGKGLKVLINKAM
ETALRQNKMKILGWVWEKNENAIATFYOKMGFVPRGAHSFYMGDEQIDMTMTKALS"
complement(200759. .201211)
/gene="BH3077"
complement(200759. .201211)
/gene="BH3077"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (MarR family)"
/protein_id="BAB06796.1"
/db_xref="GI:10175699"
/translation="MKEILREIGMIARALDSISNIEFEKVDLTGQYLYLVLRICESPG
IQEKLAEAMIKVDRTTAARAIKKLELNGFTEKRDDENKKIKKLIPTKGRKRVYPIIK
RENDHSNRVALAGFSBEEVDKMLHLLQVRKNVEVDWEFVKGNKRHY"
201321. .201959
/gene="BH3078"
201321. .201959
/gene="BH3078"
/notes="BH3078"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06797.1"
/db_xref="GI:10175700"
/translation="MKFIYVIGILFTTIGISTIQSDLGTSFDDALLVGLSTQVGLT
VGSWEILIAFMILGICNSLLKROKPEILGLVLTALIVGIDMWLFLSHFITPELWYSK
SIGFALGVVAAGTAIYLHTNIAPVDRLTUIQEIITRNLFFSRTLVLVFLFLMA
LLENGPIGVVTVTCVLCGILLNFPFPIITCKALDGLITDSRSRQNDKKHSL"
202058. .202207
/gene="BH3079"
202058. .202207
/gene="BH3079"
/notes="BH3079"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06798.1"
/db_xref="GI:10175701"
/translation="MKWFIAPFILFCAPLLALTIDQMGSTIGHIMMKFIGVGMIFFAG
LIVRS"
202341. .202448
/gene="BH3080"
202341. .202448
/gene="BH3080"
/notes="BH3080"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06799.1"
/db_xref="GI:10175702"
/translation="MAEEDFGAAFVOIGFVGSIIIFIGWILFLLVKRLKK"
complement(202804. .203154)
/gene="BH3081"
complement(202804. .203154)
/gene="BH3081"
/notes="BH3081"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06800.1"
/db_xref="GI:10175703"
/translation="MKQFFNFVTAIIFFITITLVACNNEKLADPVFNWEGENYIVTH
EPIEEELBEEIGCTVISEELPDEHGEAFKLPIGTKLFKEGKESHVIDSDSVIAVKI
DNEYTIARKISDIK"
complement(203219. .203854)
/gene="BH3082"
complement(203219. .203854)
```

```

/gene="BH3082"
/notes="BH3082"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06801.1"
/db_xref="GI:10175704"
/translation="MKNLLRAFFELTTFSEVYANESHALSQDIVENNVNNSLEDMP
ENWGKLEEVNGKVTIYEDLPDEFYVTSSEVSLNEPDPVTPSCGSCIHKKYTKVSSW
TQSSDYNGWHPGFSYGNRASGYWFSSTVSFTASVSGVSVSIAGAGSGYYNA
DYNRSRPAVHGKYIVTKYQVKEYDGAGRLIRTYIENKPSAKDTYVKIKYR"
/gene="BH3083"
/complement(204064. .204111)
/gene="BH3083"
/complement(204064. .204111)
/notes="BH3083"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06802.1"
/db_xref="GI:10175705"
/translation="MNMADFQTWVLSYPI"
204201. .204764
/gene="BH3084"
/notes="BH3084"
/notes="BH3084"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06803.1"
/db_xref="GI:10175706"
/translation="MKTAVFCSSNGASDVYKEGARQLKELARRGITLVYGASVGI
MGAVASVLGAGGEVIGVMPFLEEPEISHPLTKLIVETMHERKAKMAELADGFLA
LPGPGTLEEFEIFTWAQIGLHKQPCGLNINHYFDPLVTLHHMSNEQFLHEKYS
MALVHTDPIILLDQFSTYEPPTVKAYS"
204806. .205192
/gene="BH3085"
/notes="BH3085"
/notes="BH3085"
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/protein_id="BAB06804.1"
/db_xref="GI:10175707"
/translation="MEIRKINPENLAAPIANHYHHVAVIPRDAELVLSGIGDKNGD
LPTDIESQFVNALENIRAILHSGVDLKNLFKNLWLTEDLDKAFPSERWNEFHGYP
PTTKAFVSALARPEIKIETAAK"
complement(205367. .205735)
/gene="BH3086"
/complement(205367. .205735)
/notes="BH3086"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06805.1"
/db_xref="GI:10175708"
/translation="MGGKWFVNLLYSIKTGLNRAIEANKDPWEAEVFERHVLV
RAFSRQAHRIEGLGRKAEQQYQNGYQGVRTFWALLDSYEVMDLLEHGAIEYSRY
IYSSKGTTEEKERYFPEE"
205936. .206193
/gene="BH3087"
/notes="BH3087"
/notes="BH3087"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06806.1"
/db_xref="GI:10175709"
/translation="MEHRSLTIVYERLMLNHLVLMETKQMLQTKLAEMNGDDQSFELVKE

```

```

DLKIEEIKSYADEQSEIALDQSMPHALPNPAVTEELDTQ"
/notes="BH3088"
/notes="BH3088"
/notes="BH3088"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06807.1"
/db_xref="GI:10175710"
/translation="MKOLWEFLAQAKACLPVVFITLAIATKLADIPIPRVDIFILL
VCLLTQVMIISKLEITDELKVICVFLHGLALIEIKVHMSWSIPEEATIKFVGPL
YSGFMFASVASYMCQAWRLDLRLTNPPPTWTVILSGLIYFNFTYHYLYDIRLLK
VATLLLFRTFVFVTNHRITYSMPLSEFVLIGFFIWAENIATFFGAWPYNOEHOW
ELVKIGKISSWLLLVILVF"
complement(207184. .207807)
/gene="BH3089"
/complement(207184. .207807)
/notes="BH3089"
/codon_start=1
/transl_table=11
/product="mutator MutT protein"
/protein_id="BAB06808.1"
/db_xref="GI:10175711"
/translation="MTIEPKWLEWAKQLQIAQAAGLTYSKDVYDLERFELLREVSLEI
VAHTEVDRTIKDLFAHETGYPTPKVDIRAVVFKKLLMQVETSDGGWALPGWGD
IGLTPKEAEKVEEAGLDVKATKLIAVFDKACHPHPSPYHYKILFIOCEITGGVP
KKGIEYSAVGFFAQDELPLSLSVARNTQSOLDIAFKHRLNPDIPVLD"
complement(207937. .208398)
/gene="BH3090"
/complement(207937. .208398)
/notes="BH3090"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06809.1"
/db_xref="GI:10175712"
/translation="MRLPAYIEOLDLWEAEFRFSTSVKRVFSETDAPGHLNNTKTFVY
FEEDRTFKSLGFLKLTWNPYEGMIVADLOCDYQOROVFFDDRLTVVYKPNAGSS
SVDLHYMIKNSTGHICMTGRGTVOQVSKTKGKFFPLSEEMKDSLLNAKGAN"
complement(208563. .209318)
/gene="sdbB"
/complement(208563. .209318)
/notes="sdbB"
/EC_number="1.3.99.1"
/notes="BH3091"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase iron-sulfur protein"
/protein_id="BAB06810.1"
/db_xref="GI:10175713"
/translation="MSENVVRFITITRQDDPDSPYEETQVYPYQMNVTLSALMEIRR
NPVNSKETSTPVAMENCLVEECGASVINGKPSQCTALVDQLEQPIRLPEPKTF
PVYRDLVIDRSMFDSLKKVAAIPIDGYDLGPGPMAESKRQWAEIISKMTGCV
LEACPNNVNSKSEFIPGAPLSQVRLFNHPTGEMNKEERLQAIMGDGLANGCNSQCV
QSCPGLPITTSIAALNRSATLQSEKNTFGSTN"
complement(209315. .211084)
/gene="sdbA"
/complement(209315. .211084)
/EC_number="1.3.99.1"
/notes="sdbA"
/notes="BH3092"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase flavoprotein subunit"
/protein_id="BAB06811.1"
/db_xref="GI:10175714"
/translation="MSKGVIVVGGGLAGLMATIKVAEAGVPVPELPSLPVKRSHSVC
AOGINGAVNTKGECDSPWEHEDDTVYGGDFLANQPPVKAMCDAAFGIIHLDRMGVM
FNRTEGLDFFRFGTQHRTAFAGATTGQQLLYALDEQVRHRTSGSLVTKYGEWF
ISAILDDGICRGITAQNLKTSFQSFSDAVIMATGGPGIIFGKSTNSVINTGYAA

```

KLVEQGVYYANGFEIQIHPTAIPDDDKLRLMSARSGEGRVWYTKDKKPWFLEEKY
PAYGNLVPRDIATRIEIHVCVDLKGINGENMYVLDLSHRDKPDELDKLGLIIEIYEK
FMGDPKRPVKIPFPAVHYSGMGWIDYDOWTNPGLFAAGEDYSGHGANRLGANSL
LSATFGCWMYKRAVEYMGLEKSAESMPSSLFEGELKKEQEKDDILLKWDGENAVY
LHRELCSWTDNVTVVRYNKKLTKDEKIQIOLMERYKNIDINDTAKNSOGASFTRQL
DGMELKLARVITLGAJNRNRESGAHYKPDFPERNDEEFLKTKAKFNPTETLAEFEYEE
VDVSLIEPRKRDYSKKQGA VKS"
complement(211098. .211727)
/gene="sdhC"
complement(211098. .211727)
/gene="sdhC"
/EC.number="1.3.99.1"
/note="BH3093"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase cytochrome b-558"
/protein_id="BAB06812.1"
/db_xref="GI:10175715"
/translation="MAENREFFNRRRLSHLSGLVIPIGIFLIQHVFVNHFATAGASFNQ
AAHFMESLPFRYALEIFIIPIITYHAIYGLYIAFOAKNNTSHVGYPRNMWFLLRIS
GVFLIFPIAHVNETRIQAAGAEVNDMMADILSNPFMLVEYILGVVSATFFHANGL
WSPFVSGIYVSPRSQIMTYVVAIPFALTFFGIRALLAFINPDLANI"
complement(212035. .213384)
/gene="BH3094"
complement(212035. .213384)
/gene="BH3094"
/codon_start=1
/transl_table=11
/product="Na+-transporting ATP synthase"
/protein_id="BAB06813.1"
/db_xref="GI:10175716"
/translation="MKSALNQFMTPRMIMVSYIAMFLFSVLLFLPISQQPGVNLTF
SEALFTSVASVVTGLTVNVNSETYSYIGILFLALAIQLGSGIMTGLTFVWMLPGKK
IILSORMLNDQNGQFSGLYNLMRGILLVALSIEFIALIGLTYIRIPFVIVGAY
YQGAFAALFTNAGFDITGESLIPFADDIYVQLVTILLIFAGAIGTGPVLEVRAYS
KENPFTLSFTKITTITTFEAFVFGYIGWALEKNHYFVGLWHQQLFTFVARSATS
RSGLATIMDISLSLATLLEISGLMIIGASPSSVGGIGIRTTTLAWEFLTIRSFALGQ
DVKVFGEIHPEDQKQSFIVISVFLVLLFAAVIMISAFERGAIPLIAIFETASAFG
TCGLSMGITDELSPSQFILMVLWFIGRVGLIAFLFSVRKKEIKSYKYKPYTERIIIG"
213500. .213916
/gene="BH3095"
213500. .213916
/note="BH3095"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06814.1"
/db_xref="GI:10175717"
/translation="MSTKEQETMETVQLFGYDILLRNDVRLREILGSHDAILYWVGKS
LARKHPTETWEETFAPEKANWGOLSTLKOKROEYTFELOGPWNMTKSDORCYQLEAGF
LAQLVQWSETIAEATYVRKGSVLIDVHTDKGDSV"
complement(213985. .215223)
/gene="lysc"
complement(213985. .215223)
/gene="lysc"
/EC.number="2.7.2.4"
/note="BH3096"
/codon_start=1
/transl_table=11
/product="aspartokinase II alpha and beta subunits"
/protein_id="BAB06815.1"
/db_xref="GI:10175718"
/translation="MSIIQVKGCTSVGSVQRIEHVANRIIATAASCKKVVVVVSAMG
KTYDELGLASEITDEPRKEMDLATGEQVTIALLAMALHRKQGEATSLTQWQAGM
ITESHGNALEIVERARMDELDDKHVVIVAGFQIDQOQGEIMTLGRGSDTSVAI
AAALKARCEIYTDVFTTDPRVVKQARKLQTIISYDEMLEANLGASVILHPRAVEF
AKNVEIPLLVASSWDEPGTMIIEEGTMEQNLVVRGIAFENDVTKVTVHGLPNEIHTL
PTLFAHLSAAGINVDIIIONQDTSGMWAISEFISITQSSILKETLARLEELKGELSFGQIDY
EEELAKYSIVSGSNVSNPVAAMFEILAAINDIAIKMWSTSEIKVSAVIDEELMWLVAV
EMLHQPKLHEEPEPAVAVKV"
complement(215611. .217386)
/gene="uvrC"

complement(215611. .217386)
/gene="uvrC"
/note="BH3097"
/codon_start=1
/transl_table=11
/product="excinuclease ABC (subunit C)"
/protein_id="BAB06816.1"
/db_xref="GI:10175719"
/translation="MKDHIKQKLTVPDQPCVYLMKDRQGTIIVGKAKVYLNKRVRSY
FTGSHDAKTORLYGEIADFEYIVTSSNIEALILEMNLIIKKDPKYNMVLMKDDKSYPL
KLNEEHPRLVTRTKLKKGGKFGPYPNAGANETKKLLDRLYPLKRCRTLDPKVCVL
YIHIGQCLAPCVYEVTPEQNEQMVQEIITREFLYGKHVKVELEKMKHAKAEELNFERA
KELRDTLGYMEAYMEKOTMNDVRDVFEGYAYDKGCMCVQVFFVQGGKLIERDVS I
FFPYKAEEDLTFIGQFYLOKDHIKPGEVLLPAGTDAALVLDVSVHPIKRGKKK
ELVDVARNATMALKEKFALIERNEERTIRAVEQLGEAMGIPTPYRTEAFDNSNIQCT
DPVSAMVAFIDGKPNKEYKYKIKVTVPDDYESMEVVRVRRVRLLEKQSLIPDLI
VIDGKGQIAAQAELHDELGSLSPVCGLANDEKHRTSQLLGLDPPPOVPLKRDSSHFE
YLLQRTQDEVHRFALTFRHQRTRSKTFQFQSVLLDDVPGIGEKRRKQLLKHFSGVKMKREA
SIDDFALSLPKSVAETLYNKLQTK"
complement(217510. .217824)
/gene="trxA"
complement(217510. .217824)
/gene="trxA"
/note="trxA"
/codon_start=1
/transl_table=11
/product="thioredoxin"
/protein_id="BAB06817.1"
/db_xref="GI:10175720"
/translation="MAIVNVTDDTFAQETSEGVLADFWAPWGPCPKMIAPIVLELDG
EMDGKVKIAKLDVDENQETASKFNVMISIPTLIIVFKDQVYDQVTGFQPKDALAEALLNK
HL"
complement(218073. .219050)
/gene="etfA"
complement(218073. .219050)
/note="etfA"
/codon_start=1
/transl_table=11
/product="electron transfer flavoprotein (alpha subunit)"
/protein_id="BAB06818.1"
/db_xref="GI:10175721"
/translation="MARKVLVIGEVDRGELRNVSFEEAAGKEVANGEVGVLLGLEN
ISDLANOLIHYGADRVVAVNHPDLAQYTDAYQAIQLVLDVSEPAALILGHTSIGKD
VAPRVAMKLGAGLISDVISVEKDGDEIVFTRPITYSGKAFKKEGIIKFIATIRPNNI
APLEODASKSGDVSSLTVDIKDLRTTREVVRKATGVDLSEAKVIIVAGRGVKSKEG
FEPQLKADVLGAAVAGSARGACDADYCDYSLQIGQTKGYVTPDLYFAIGISGAIQHLA
GMSNSKVIVAINKDPEAPIFEVADYIGVLDLFEVVPVLLTTELKNVLVSS"
complement(219065. .219835)
/gene="etfB"
complement(219065. .219835)
/gene="etfB"
/note="BH3100"
/codon_start=1
/transl_table=11
/product="electron transfer flavoprotein (beta subunit)"
/protein_id="BAB06819.1"
/db_xref="GI:10175722"
/translation="MNIIVLMKRTFTEEKISSGQIDDSGAEFIINPIDEVAIEBA
IVLRDKHGEVTVVVGEEAEKEKELTALAMGADKAVLIDSELDLDQYTTATLLAA
YLKQDEFDIIILGNVAVDGGSGQVAPRVALLGTPQVTTITSLDIEDKRATIVRDVEG
DEETVAASLPVLTAAOQGLNEPRYPSLPGIMKAKKPLETLDLDDLDLEEDVEAKTT
TLQVFLPPEKSGACKVLEGEVAQVKELVSLKSEAKVI"
complement(219857. .220633)
/gene="BH3101"
complement(219857. .220633)
/gene="BH3101"
/codon_start=1
/transl_table=11
/product="enoyl-CoA hydratase(3-hydroxybutyryl-CoA
dehydratase)"
/protein_id="BAB06820.1"
/db_xref="GI:10175723"

| | | | |
|------|--|--|--|
| gene | /transl_table=11 | | |
| | /product="phenylalanyl-tRNA synthetase beta subunit" | | |
| | /protein_id="BAB06829.1" | | |
| | /db_xref="GI:10175732" | | |
| | /translation="MLVSYOWLKQYVDLGGISAEAEVKLTRGGIEVDIIDLHNOGATGVVGHVLECRHPNADKLNCQVEIGEEEPVOIVCCAPNVAAGQKAVAKVAGVAGLGNFKIKKAKKGLSELQSGLCSLQELGVENKFKQREFADGIIYFSDVDEIGSDALEARNLDYQVLELDLTPNRSDCLNMGVATVEAVALDEKVEHWPTESSGTDGQASDYVSEVEDIEKRPYVGATGIKINVKYGPSPWLQNLRIAGIRPISNVVDVNTNVLLEYGQPLHFVQYDRKSNKVTFLAKEGEKQWMTLDVVERTLQAOHLHLLTNGRDVPVATAGVMGGATSEVQEDTITVLEAAYFPDPAVKRASRDGLGRSDSSRSFEKGVDPKRVREAAARAAATLIAEAGTVVGVGVPAAGLATEEPPVSLNLPHMNORLGMELSHDEVAALIDRLGFAFTVENDTFEVTASRRGDRIEDLEFEVARLYGYDNITPTTFVGGTTRGGULTDYQKRRKLRRLVLEGALFEWNTYSLTSEEKKGQGGENEDYLPDIRVAMPSEERSTMTSLVSHLLHDVYQGNLNRKQEAFLTIELGSVFLSKEETLTQKPEERMLAGALTGFWTEHLWQGEIKRQDDFFVWVGLGEGMFAELGLSEKVTPERGEQPMHPGRTANVKVNGRAVGFAGQIHKPVSSELDKETYVPQLDVDESLLTEEGFEVAYEPLPRFPAISRDIALVVDENVTAALQLOQIVFANGGEWLKHVYFDLVEGEHMEAGKKSIAFSITYFDPERTLDEEVTAVHEQILKLEASTGALVRG" | | |
| | complement(234269..235303) | | |
| | /gene="phes" | | |
| | complement(234269..235303) | | |
| | /gene="phes" | | |
| | /EC_number="6.1.1.20" | | |
| | /note="BH3111" | | |
| CDS | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="phenylalanyl-tRNA synthetase alpha subunit" | | |
| | /protein_id="BAB06830.1" | | |
| | /db_xref="GI:10175733" | | |
| | /translation="MRQLEALRDADTKKIEQAQEKVKQVQIEIRVKYLGKGKPVTEVLRGMGLSPTEPERVVICOLANDRGAETKAIKSMKALEQAEIHRKLQEETIDVTLPGRPVLTGGAAHLPTAITITVEDTFTGLGSVAEGPQIETDYNFEALNLPKDPHADQMDEGVYFKDELLLRTPSTVQOATMKYKGGPIKIIICPGKVFRRDDDDATSHQPMQMGTEGVLDHGVMSDLKGVLAQAFKSPFGEERSIRLPSFPETPEPSVEDVSCGICGCGDGRVCQTQVIELLGGWVHPRVLEMSGFNPNSEYSGFAGMGVERLSMLKGYGIDDIRHFYFYNRRFLAQAKRV" | | |
| | complement(235732..236487) | | |
| | /gene="BH3112" | | |
| | complement(235732..236487) | | |
| | /gene="BH3112" | | |
| gene | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="rRNA methylase" | | |
| | /protein_id="BAB06831.1" | | |
| | /db_xref="GI:10175734" | | |
| | /translation="MKOIDSPKELVKKKLHKKGRDSQGQFLIEGYHLVEALKSTITITELVYESVLTPTAHVDERDVLISDRVWADLTETKTPQGTAAICHIIDPEPSFQELGEMFLICIDHVDQDPNGVMTIRTCDAAGSAVIVSEGVSDIYNKSVRASQGSIFHLPVILKGLHWVDQFKEAGIPVFGTALEGGSTVTAIEPQARFALLVNGEGEGVEPELLDKAQNLVILPLYGKAESLNVAVATGILLYLRD" | | |
| | 236604..236810 | | |
| | /gene="BH3113" | | |
| | 236604..236810 | | |
| | /gene="BH3113" | | |
| CDS | /note="BH3113" | | |
| | unknown conserved protein in B. subtilis" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /protein_id="BAB06832.1" | | |
| | /db_xref="GI:10175735" | | |
| | /translation="MNFNLRAIMANIQQSNEQEVEATIVDAIQSGEEKMLPGLGLVLFVYWKNCNEQKDLQCEIQISGLQ" | | |
| | 237028..237891 | | |
| | /gene="BH3114" | | |
| | 237028..237891 | | |
| gene | /note="BH3114" | | |
| | unknown" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /protein_id="BAB06833.1" | | |
| | /db_xref="GI:10175736" | | |
| | /translation="MNFNLRAIMANIQQSNEQEVEATIVDAIQSGEEKMLPGLGLVLFVYWKNCNEQKDLQCEIQISGLQ" | | |
| | 237028..237891 | | |
| | /gene="BH3114" | | |
| | 237028..237891 | | |
| CDS | /note="BH3114" | | |
| | unknown" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /protein_id="BAB06833.1" | | |
| | /db_xref="GI:10175736" | | |
| | /translation="MNFNLRAIMANIQQSNEQEVEATIVDAIQSGEEKMLPGLGLVLFVYWKNCNEQKDLQCEIQISGLQ" | | |

| | |
|------|--|
| gene | translation="MSPFAPSELSCYLSDMGQRDLPLKTNALHYKEIVSRKKQLLE TKLGTEFLFQOQVITISGFVYVTTNOLSSNPISFSATKELLISAPFKEVTLTFSVKQ CRCBALCFQDEQEQIKYIINIDITDAEKQKILLSKGTSAPFKMKMCIETNNI FDSVSYQKIVTFNPKMDSSTSHGSHIVQHSEPAKVFRAVDPPFAPSQIOIENHS DCPMKAIVHVDYERKVSQIIIEKYQKVLDFVSEIQRLVIKGDGEGGVCEGAFSLFLSH KKPVSRSDDL" |
| CDS | complement(237910..238419) /gene="dinB" complement(237910..238419) /gene="dinB" /note="BH3115" /codon_start=1 /transl_table=11 /product="nuclease inhibitor" /protein_id="BAB06834.1" /db_xref="GI:10175737" /translation="MSHAKTLVQYNWANGRIMEHLQITISPEKYEOEIKSVFPSLAK VCSHLVLEYLWLDILSGDKGVAALEEVFQLOHETDALPLEEAAAYRTLSERFLSL EKENLERRLLNNPYVGERETSUAEIVLHAANHGTTHRGNISAMHLQGDGASVMTDF VFYMYQEDL" |
| gene | complement(238747..239664) /gene="BH3116" complement(238747..239664) /gene="BH3116" /note="BH3116" unknown conserved protein in others" |
| CDS | complement(238747..239664) /gene="BH3116" /note="BH3116" /codon_start=1 /transl_table=11 /protein_id="BAB06835.1" /db_xref="GI:10175738" /translation="MPKRYQNLDEVTNKLIVELWKLCLRERWSKSDLSIHPRADEP NVEELKNRTKTSYTWLGHSLFGLQMKGLNILDVPAKYLGFSLKSLISQILSDEL EIDVTWISHGHDYELTSLRSLGEPFVYPAGLNLLFKQKGLHRRVEEANWMDTQFA QDLTIHVPQAQHWTRRSLLFTDNTSHWGQWYLESDEETNVEVGTGYPGRQEGIGKQFA LDTVMPTGSVEPEMFTYASHINPEDAKAFVLOAKTFVPMHYGTIVRLADDTGPEAL ARLHEWKRCOLPEHQFLVWKVGTETFKK" complement(239829..240521) /gene="BH3117" complement(239829..240521) /gene="BH3117" /codon_start=1 /transl_table=11 /product="RNA polymerase ECF-type sigma factor" /protein_id="BAB06836.1" /db_xref="GI:10175739" /translation="MNDKHLKQLHDTVSRVCLFTVTKSTDWDAEDLAQETWUKAIELTK KQGHQNEPALLRYAKNQWIDQSRKKAVAFILKQEKPEVTPPEQGLFEMERAPFAVL NHLSPQLSVFLLRDVFGLGYSTSEAKWVEKTEGAARLAKRRKRSKSLALVRDDLETENS PLPNGEAKWLRAIVAYEAGEITVIDLIDQGVYAARTMVRVSGQERKRVAVLPDNGAF RRSDPTGCLAA" 240675..241259 /gene="BH3118" 240675..241259 /gene="BH3118" /codon_start=1 /transl_table=11 /product="ATP-dependent Clp protease proteolytic subunit" /protein_id="BAB06837.1" /db_xref="GI:10175740" /translation="MNGVYKVPVIEQSGLSERSYDIYRLLKDRIVFGAIDQOJANS IVALLFEANPKKDIYLVINSPGSGTSAGFAIYDMQVKPSIHPTICGTGMAASFAA ILLAGTQKGRFALPNSEIMTHQPSGAQOASDLATAKRIILIREKLVTVTSRTGT QSPKRVAKMDRDYFMSAEAELEGVIVDQIITSV" complement(241369..241773) /gene="BH3119" complement(241369..241773) /gene="BH3119" /note="BH3119" unknown conserved protein in others" |
| gene | complement(241369..241773) /gene="BH3119" complement(241369..241773) /gene="BH3119" /note="BH3119" unknown conserved protein in others" |
| CDS | complement(241369..241773) /gene="BH3119" /note="BH3119" /codon_start=1 /transl_table=11 /protein_id="BAB06838.1" /db_xref="GI:10175741" /translation="MNGVYKVPVIEQSGLSERSYDIYRLLKDRIVFGAIDQOJANS IVALLFEANPKKDIYLVINSPGSGTSAGFAIYDMQVKPSIHPTICGTGMAASFAA ILLAGTQKGRFALPNSEIMTHQPSGAQOASDLATAKRIILIREKLVTVTSRTGT QSPKRVAKMDRDYFMSAEAELEGVIVDQIITSV" complement(241369..241773) /gene="BH3119" complement(241369..241773) /gene="BH3119" /note="BH3119" unknown conserved protein in others" |

/translation="MTVKAQHIFVNLVKDLNKTKEFTTSIGTFEFNEQYSDDKAACLV
IGENIYAMLLLEQFFQSFNQPTVDITSETSRVIAIAISDRSQVQVDEIVTKAIEAGGKP
ANDPMDGFMYSWSFQVDVGHLEVLVMEQKN"
242027. .242518
/gene="BH3120"
CDS
242027. .242518
/gene="BH3120"
/note="BH3120"
unknown
/codon_start=1
/transl_table=1
/protein_id="BAB06839.1"
/db_xref="GI:10175742"
/translation="MTMKISPEMTVEEVGAALQELDSRWQFVWEKHDELQKLYPEH
GDATYGYMFKLLPPIWEQVEQDFYSALPEKDDYLGGLCNFRHSMKEHMGSPGH
NIRVFWIVLANQHDHVGSLLLEHHSHERFLPAPPRICQETIREKTIHAIRQLQ
EOV"
complement(242817. .244013)
/gene="BH3121"
CDS
complement(242817. .244013)
/gene="BH3121"
/note="BH3121"
unknown conserved protein
/codon_start=1
/transl_table=1
/protein_id="BAB06840.1"
/db_xref="GI:10175743"
/translation="MSIFENIRMLHSHIKAKMRSVLTMLGIVIGVAAVLIIVVAGQG
GEALLKSQJAGEENTVDIYSPSEELIQANPNWQAPFSQEDI RELEK IDEVEHVVA
FSEYASVRQEOQTDASIGVQAYNLVYPFNVTGVAIHTGDIILGRRVALISEGI
AEELFDEVSPIQGVIRIGNQPVETIIGLEPPTGLFAFNQVEVYIPWE"WRNYIGKSD
VPTTKSDLENLQJAGEKATEILNARNGTEAYOVLANNEEIAEGIGQITRMTMI
GGTAGISLFGVGGIGVNMILVSVTERTREIGIRKSLGATROQILIOFLNEAVILTIG
GILGIIIGSGIATLVSTIAGWSLSISVQYIVGGLLFESMLIGVIFGLLPANRAAKLANPI
ESLRYS"
complement(244016. .244699)
/gene="BH3122"
CDS
complement(244016. .244699)
/gene="BH3122"
/codon_start=1
/transl_table=1
/product="ABC transporter (ATP-binding protein)"
/protein_id="BAB06841.1"
/db_xref="GI:10175744"
/translation="MIKLERTKSFVRGTEVWEILSAIDLEIASGDFLAIMPGSGGK
STILNIGLIGDQPTSGRYMFDKDLTNYSEQETAKIRNHHIGFVQOQFHLLRLRLAQ
NVELPMVYAGMKKERTERAHALERVLGAERTYLPNSLSGGQKORVAIAKSI VNEP
NILADEPFGALDTKTSITETIMELLCSLNNEGFTIALVTHEPEIAETOOTVFRDQOI
VETVRNGGD"
complement(244692. .245816)
/gene="BH3123"
CDS
complement(244692. .245816)
/gene="BH3123"
/note="BH3123"
unknown conserved protein
/codon_start=1
/transl_table=1
/protein_id="BAB06842.1"
/db_xref="GI:10175745"
/translation="MNKKLLISLALVIGIGAVATIGVNMNGESAASVEVVSPEMET
ISAEIWPCTVELVDMQRYADASLGNDYIELVEYDVEEGTPLIETYSRQLETTERS
QLEUQIESSYLRINLHKEQKELQKKEKELRDEVGKEADKTIOPEKDQDFEKRAN
LDRLQLOIKQDLEQKQKDLVVKSRGDCGTVMVDHLOOQGDGWTKPVLEIASAGELQ
LQKTKISEQVLSVISEQVRISDSDALPEEWEGTVKYSYFPSEAEITLGNAGIVOYP
ITVOLSGDQSKLKPGYOMIMHIVTEKTALIVPLEAVHQGDGGSYVYVVEDGKARQVE
VETGLVENSREVRIGLREETDQVIVSDGVLTDGMDVWSVD"
complement(245827. .246501)
/gene="BH3124"
CDS
complement(245827. .246501)
/gene="BH3124"
/note="BH3124"
unknown conserved protein in B. subtilis
/codon_start=1

/transl_table=1
/protein_id="BAB06843.1"
/db_xref="GI:10175746"
/translation="MSVNNETMEHENAGGKPSIIGTSPVLQERIKONPKVLVPL
LLIFMGIGLVSLVATLVELVPPPEIAAMPPEGRVFRITLIIGIFMILLVPL
ISATVOLLIAKAAKDVTFKOLFNMNLFVGISLAGLLNSVIAFLFGLSLLSTSTSL
GTFEVMDHALFPLLNTIEIFTWGVIIAFLQKQVAGFGKLASWIVTIFEVIMILLS
LTGIGV"
246836. .247264
/gene="BH3125"
CDS
246836. .247264
/gene="BH3125"
/note="BH3125"
unknown conserved protein
/codon_start=1
/transl_table=1
/protein_id="BAB06844.1"
/db_xref="GI:10175747"
/translation="MKVSSKGCEYCLRALIILAKOPERRFTVHLSKHTLPPPPYLEKL
LSQRLVLSYITSKRKGKGGYQLSRDARDISMGVIRQLEGPAPMCSVSVTSYQSCEL
EQGCLLKPLMGLVRDVTVANVLDQTTLADLIENKLGQRKV"
247265. .247468
/gene="BH3126"
CDS
247265. .247468
/gene="BH3126"
/note="BH3126"
unknown conserved protein in B. subtilis
/codon_start=1
/transl_table=1
/protein_id="BAB06845.1"
/db_xref="GI:10175748"
/translation="MAYHDPKEIEIKKAIKIKYGSVLTIHEDQIMQIDATERVRF
PQNKQKQPKKSGSQPRTSPSRN"
247537. .248604
/gene="BH3127"
CDS
247537. .248604
/gene="BH3127"
/codon_start=1
/transl_table=1
/product="sulfate ABC transporter (sulfate-binding
protein)"
/protein_id="BAB06846.1"
/db_xref="GI:10175749"
/translation="MKSVIQKSOAYFKGITILTMIFALTACSTDSTSNQEQSSDTK
AQPIELLANSYDPTRELYQDFNRSFIEYQOETGETVTIQQSHGSGSOARSIIDGLE
ADVATLALADLMDLHETRNLI PENWOERLPYNSTPTSTIVFLRQGNPLDIQDWD
LVQEGYSVITTPNPKTSGGARWVLAAGVALQOYGSSEAKKEFVKLYONVAVLDSGA
RCATTTFERGIGDVFSLWENAEFLAHLGELGDQDQDIPFPSVSI LAEPPVTVIDANVD
KKTREVAEAYLEYLYTDIGQELAAEHYKPREEDILAKYSEVFPETIQLFTIDEVYGS
WSEAQKTHFDGCGVFDQIYQP"
248655. .249494
/gene="cyst"
CDS
248655. .249494
/gene="cyst"
/note="BH3128"
/codon_start=1
/transl_table=1
/product="sulfate ABC transporter (permease)"
/protein_id="BAB06847.1"
/db_xref="GI:10175750"
/translation="MKSYSRKNHNLPCFGLSLGFTMMYLGILVLLPLSMVFINFSS
MGWQAFWQATIEPRVLASIRLSFGAAIIAASINAVFGLLIAMVLVRYHPGKRIIDGL
VDLPALPATAVGTALTYTTNGVIGQYLEVIGRIATPLGTVIALTFIGLPEWR
MVQPVLOGIETKELEBASACLANRLQTSKIIFPTVLPALLTGAFALAFALGEYGSV
VFISQNLPMQETITPLLTMTKLEQFDYAGATAALAAVMLIISFFMLLFLINILQWNSQRR
QLS"
249507. .250370
/gene="cysw"
CDS
249507. .250370
/gene="cysw"
/note="BH3129"
/codon_start=1
/transl_table=1


```
/product="sulfate ABC transporter (permease)"
/protein_id="BAB06848.1"
/db_xref="GI:10175751"
/translacion="MAQLSTNTPIFOOTAANESREVKVTLIGTALSFLFLFLVPLPIA
IFYRAEFOGLAVFASISEPDALSAIKLTLFLVLIIVPLNTLFGIAAAWAIGFQERG
KNLITFIDLPFSVSPVIAGLIFVLIFSFGVFGPMLLDGRGIQIIFISPIGIVLATFV
TVFPAARELIPVQAQTAQEASLTIGASGKTLNHWTLVLPNKWALLYIGVILCNARA
IGFNAVSVYSGHRIGLTNTMPLHIEILYNEYOFTAAFAVASLSLIAIFTLIVKNI
EMKSTVRRGS"
gene
250370..251443
/gene="cysA"
CDS
250370..251443
/gene="cysA"
/notes="BH3130"
/codon_start=1
/transl_table=11
/product="sulfate ABC transporter (ATP-binding protein)"
/protein_id="BAB06849.1"
/db_xref="GI:10175752"
/translacion="MSIVIQNVSKSFGSFOALADINLSIETGELVALLGSPSGKTSL
LRIIAGLEAADOQDIYFKDKVQTHAASRQGVGFVFOHYALFPHMTVADNISYGLRVK
PRKERSKEIAEKVRELLALVKLEGMDDRYPAOLSGQQRORALAPAVLAVERKVLIL
DEPGALDAVKRDLRKWLKLNREQVTSVFTVHDOEALDVSDRVVMNQCKIEQV
GSPDEYEQSPFVDFLGNVNVFTGRVKQGVQFDKHLQTKPTISKDTHDQAVVY
TRPHHMSIKRKORDAIPAVIEHIMVGPITAFLELNWKKDDEEVLQTELAKDRFHELDL
KKGTVFVVPKNLTLPFPEEFTI"
gene
complement(251524..251976)
/gene="BH3131"
CDS
complement(251524..251976)
/gene="BH3131"
/notes="BH3131"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06850.1"
/db_xref="GI:10175753"
/translacion="MKTPLPLTDEERKSRRQVKVLSDLADLEVEALATYLHSTVTR
ARYLRGAIFQTLISIGPKVAQMVVDLGYDFLEAVKFEKGDGMVDRLEKLYGWVDFC
VEDCFRIVHANDRESEKHWDFTTERRKAYRATSGYPSDRPTQSWYA"
gene
complement(252136..253221)
/gene="BH3132"
CDS
complement(252136..253221)
/gene="BH3132"
/codon_start=1
/product="endo-1,4-beta-glucanase"
/protein_id="BAB06851.1"
/db_xref="GI:10175754"
/translacion="MAQLDERLQMLKELTDANGIPGNREARDVMRKHIEPFADEVTT
DKLSLVAKKVGKEDGPKIMVAGHLDEIGPMVTSIDDKGFLRFQTGWGWEQVMAQR
VTVMTKGNIPGVIGSKPPHILPAEARKPVDDKDMFIDIGVASKAEALFEGVRPGDS
IVPVCDPTVMNEKYLMAKAWDNRIACAIAIDLVLRELKGTNHPNVVYGVGTQVEEVL
RGATSNHYIKPDIGCVGVGIAGDTPGVTEKEALAKMGEGPOLVLYDASNVSHKGLR
DFVTDIADENIPYQDAMAGGDTSGAIIHLTANGVPALSTVATRYIHSASILHRD
DYENTVKLIVEIKLDRDVTNRTFD"
gene
complement(253336..253824)
/gene="BH3133"
CDS
complement(253336..253824)
/gene="BH3133"
/notes="BH3133"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06852.1"
/db_xref="GI:10175755"
/translacion="MNLNKLFTLQOTLNERIQTEHGLTGVDLFHEQLLALKVELGELA
NETRCFKYWGKGPSTTVILEEYVDGLHFVLTLGLTLMRTDVEIKQMDHETNLTLEQ
FLLVFRYISELEQSRTEQAPRTLFSHYIALGEKLNFTVEIEKAYLIDKNKVNQRQDE
GY"
gene
253994..254404
/gene="BH3134"
CDS
253994..254404
/gene="BH3134"
```

```
/note="BH3134"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06853.1"
/db_xref="GI:10175756"
/translacion="MVIARLLILIALIILIIYSAIRYLFHPKRLKEQAHRHQQHFFVFD
DEKVRNFLTLYKGVMEGEKYLGTETFEIVSYVWARSTDLKGLERQDFRFE
EETILRYPTARIEMKSPVKEFLAKTKDEQPTPSP"
gene
complement(254444..255010)
/gene="BH3135"
CDS
complement(254444..255010)
/gene="BH3135"
/notes="BH3135"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06854.1"
/db_xref="GI:10175757"
/translacion="MEEMFETSLKVIQSGWLAPVLFIMLHVFRQVFFIPVIIICLVG
GYLFGSYGSLYSIIGLTATSAFYGIHVGFPRFLAKVSQSKRFLGEEKESLSLQPM
LLRVIPFIHPLVSLYVMETRPTFREYMKISFVLSMPPAIIYITLFGDMIHELPLVGV
VFALFGLCLCFMARKEQRYAWQEFKA"
gene
complement(255039..255323)
/gene="BH3136"
CDS
complement(255039..255323)
/gene="BH3136"
/notes="BH3136"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06855.1"
/db_xref="GI:10175758"
/translacion="MGFELVFFLYVGLTVIGFVAVMGLDKOKARTGRWRTPEKTLWL
IALLGSIGVFCAGMYFRHKTCHASEVGLLIVQLMILAYIFYDTRS"
gene
complement(255438..256655)
/gene="BH3137"
CDS
complement(255438..256655)
/gene="BH3137"
/codon_start=1
/transl_table=11
/product="transposase {01}"
/protein_id="BAB06856.1"
/db_xref="GI:10175759"
/translacion="MNETMDPLGLKGVKVTMEEREGFIALHVEMPRTPHRCPSGGER
TERTHYRIOKVOHLKLFERWVLYFYKRYACRCGRFAEKACFERYORQTVENNQ
AFRLRIQGNFTDTANQFQTSVTTAMRFDQLAAPLYKKVDHLPPRIADDDYKGTN
AGYQVLIADADTRKPLDILDRKADTLKRYLQEKHHVKTVMVMDLSYTFKSAVQAK
GHPLIITGDPFPHCRYIYWALEGVRRRYQKAFHPYHRRKRIKHLFOKERLDDQK
KKVERYLESPDLREAYELKEHFRWYQEAQMQRGEMSEVQGLYAFYRRYEAASRL
KEFHRAIQTUKNQTEILNSFAEGLNNGFTIEGLNQTQVKIKRNAFGPRYDRLLNRIL
LHQFKHQTFGVG"
gene
complement(256818..257177)
/gene="rplT"
CDS
complement(256818..257177)
/gene="rplT"
/notes="BH3138"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L20"
/protein_id="BAB06857.1"
/db_xref="GI:10175760"
/translacion="MPRVKGYVARRRRKKVLAKGYFGSKHTLFKSAQGVQVMSIQ
YAYDRROKKRDFRKLWITRINAARLNGLSYSLHGLLAGLAINVRNKRMLADLAIN
EKAFQAUEKAKASLNN"
gene
complement(257211..257411)
/gene="rpmI"
CDS
complement(257211..257411)
/gene="rpmI"
/notes="BH3139"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L35"
```

```

/protein_id="BAB06858.1"
/db_xref="GI:10175761"
/translation="MPKMKTHGGAARFKKTKGSKLKRSHAFTHLFAFKSQOKRKL
RKSAlVHSGDFKRIREMLTYKK"
complement(257449..258021)
/gene="infic"
/gene="infic"
complement(257449..258021)
/gene="infic"
/note="BH3140"
/codon_start=1
/transl_table=11
/product="translation initiation factor IF-3"
/protein_id="BAB06859.1"
/db_xref="GI:10175762"
/translation="MALFHALISSWKMLRWLIISKMLVNEGIRAREVRLIGANGDQ
IGVKSAEALDMARNVLDLCVAPNAPKPPVIMDYGKYREQOKDKKARKKQKI
NVKEVRLSPITIEDHFNFKLRNARKFLSKGDKVKAAIRFRGRAITHSEIGRNVLERLA
KECEDIAIVEAKPKMEGRSMFLVLPKTR"
complement(258289..260226)
/gene="thrS"
complement(258289..260226)
/gene="thrS"
/EC_number="6.1.1.3"
/note="BH3141"
/codon_start=1
/transl_table=11
/product="threonyl-tRNA synthetase 1"
/protein_id="BAB06860.1"
/db_xref="GI:10175763"
/translation="MADMINITPDGAVKEFPKGTTHAETAGTSPGLKKKALAGMLD
GTLLDNLPTIEODGTITIVTPEDEALEVLHSTAHVMAQALRLFKDRNVKLGVPV
IEGGFYDVMDSESLPEDLPKIEKMKKIIGENLPIERVVVGREALARYEEVGDPY
KIELNDLPDEDTITIEQGEFFDLGRVHVPTGKLFKLNLAGAYWRGDSNNKM
LQIYGTAFKADLDEHLEAKERDHRKLGKELGIFALSKYGGGLPLPKGA
TIRRIERYIVDKKEGLQYHVTPVLASSELYKTSGHWDHYKDDMPFTWMENEELV
LRPMCPHHMVKYTEMRSYROLPIRAELGLMHRVEMSGAVSGLORVCGTFLNDAHI
FCRPDQIKDFEVRVRLIQAVYEDFGLNKYSPLSRDPDEKRYFDDDNWNAQOM
LEKAMELELYFEAGEGAAFYKPLDVQVTKALGKDETLSTVOLDFLLPERFDLTV
KNQVETLQAQIRVEDERDEKIGYKIREAQMKIPLYMLVGDKVEANGVNRKYGE
KDSMSGLDFVRHVARAKK"
complement(260646..261512)
/gene="BH3142"
complement(260646..261512)
/gene="BH3142"
/note="BH3142"
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB06861.1"
/db_xref="GI:10175764"
/translation="MISIHFEKAHDCQLFQQLQGYLTKYAELGLKGGIRAGGDETIY
VNYENQVNFYDSFOLLASIFTDYVISTKEEWLDDIENMEFYTDEEKEQIVAI
RSLGLEDDMPSLMFTNRREYIYQAFSKHIEENMTFYDPFLTRLEKYGELLIDC
VELAIDYEMEQEYQSMVETLROFIHSRIPKRGVIYAVDERFHFDEHFRLLSTDEI
VYDLKELVFEHGLSVEEMVISPLVSMLPETVKVFTNDEEHGVILLQLAIFQRLKVF
PLKDSPHQREE"
complement(262084..263196)
/gene="BH3143"
complement(262084..263196)
/gene="BH3143"
/note="BH3143"
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/protein_id="BAB06862.1"
/db_xref="GI:10175765"
/translation="MSIDGILERVNGERLSMEDAVKLVESDEVEKMGAAANOIMLKW
HBPITTFIGNRVNNTFCDTYRCFAPYRAPHGKEGVYLDVDEVLKIKIETIDYGG
TELLMQGNGNPDLTIDYTDLLRNKIERFPNIMWHSFSAEYWKIAEVSMSVEEVL
ELHEAGLDSMPGGAEILTEELTRVSLRKITWEQWINAMKATKVMGHGTATMVG
GSFEERLALQVRDAODETCFTAFISWLFOPTNGMYTKKLPDYLKNAVLSR
LFILDNIPNQSSWNTMGPEVGKLSLQYGCNDFGSTMIENNVSAAGTTHKVTNKILQ

```

```

LIREAGKIPAOQTTSYEIIRTFEDKEAAEKDFVMQN"
complement(263258..264193)
/gene="dnai"
complement(263258..264193)
/gene="dnai"
/note="BH3144"
/codon_start=1
/transl_table=11
/product="primosome component (helicase loader)"
/protein_id="BAB06863.1"
/db_xref="GI:10175766"
/translation="MESIQSSLKHMTGNRSPEKQFTOLKEAVFRSPHVOLFLEHPSL
SPITLEQGLSKLYEQKESHAHCGLQKCPNLMKGYQPTLYVERDSLELSYSPCPL
KEEREKKRSIRLSYIPKEILEAKFDDVSEPGRSIASHRALEFALSAPGEDGM
GLYLGFGVGTFLMGAIANELKDRGIDSTIVVVPDFRELKOSIGDGTFOOKLDFV
KNAQVLIFDDIGAETMTSWRDDVLGVLQIYRIMEKLTPTFTSNYDYDELEHLYAND
KSGTELLKARVMERIRHYTVSNVWQGNRRHQSS"
complement(264194..265651)
/gene="dnab"
complement(264194..265651)
/gene="dnab"
/note="BH3145"
/codon_start=1
/transl_table=11
/product="chromosome replication initiation/membrane
attachment protein"
/protein_id="BAB06864.1"
/db_xref="GI:10175767"
/translation="MTWHKELLPVDRYVMVRYADYLVHEVHQKVLTLIYQPLVGSINYS
LYMTWSELARDDYWSEEQTHRQLMGMVDLPAIFORKKLEAIGLLRTFKTKADGV
TTYLEQLQPPMPQOQFENDVLSVILFKLGKQYRKLRERFVVDVDTFTTETYG
FDEYFTLSHSEISVSHSEMSVDSLALYHQQFVSEKSGSFHFKSFDFELKDKL
SSFTVEPHTTTPVKETILRLAFVYRTEPLQMSILQOVLVHDDVDIGELRKRTOEW
KLEHNEPHTLAVKOTHLPLAKRTMANOQPOTEEEKVIALYETTPPLTLLESRSPGKV
PLADLKIVESLIVDYQLPGVNVNLLYLLRNDMKLSKSLIEKLAGHWARKKVKYTV
EAMTLAKEERKLETWKKQDSAKTSGKTVKKNRRDQLPLWLEEERKLQEQQPAV
RENLGASTMDLDEERRFERMEERQORQWKEEG"
complement(265710..266171)
/gene="BH3146"
complement(265710..266171)
/gene="BH3146"
/note="BH3146"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06865.1"
/db_xref="GI:10175768"
/translation="MRCPACHNGNTRVLSRPAHEGRSIRRRRECESCNHRFTTFEMI
EVPFLIVVKKDGRQEFSSDKILRGLIRACEKRPVPLETLEGIVNEVERELRGOGRNE
VDSKEIGELVMERLANVDDVAYVRFASVYRQFDINVFIOELKELMERDDR"
complement(266336..266737)
/gene="BH3147"
complement(266336..266737)
/gene="BH3147"
/note="BH3147"
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB06866.1"
/db_xref="GI:10175769"
/translation="MGFRFETLSQVSRRTTETHHNSNPSTQTHYKTKDKATAAVER
VMOQSGFTVKRVEQERGEVIAQTSQKSLLVATIVMKPFTAVDFSCSTDTILPDS
FGHSKRLILSLYEQDLKELPYIGSGDEL"
complement(266916..267299)
/gene="BH3148"
complement(266916..267299)
/gene="BH3148"
complement(266916..267299)
/gene="BH3148"
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB06867.1"
/db_xref="GI:10175770"

```

| | | |
|------|--|--|
| gene | /translation="MDTMGRHVIAELWGCDVDKLDLDSFIEQVFDVDAALKAGAEVREV AFHKFAQGGISGVVIVISESLTIHSFPFHGVSASDVYTCGDRIDPNVASNYIAEALKA TATEVFLPRGMPIQIEKPKVKVL" /db_xref="GI:10175771" | |
| | complement(267591..268625) /gene="gapB" /EC_number="1.2.1.12" /note="BH3149" /codon_start=1 /transl_table=11 /product="glyceraldehyde-3-phosphate dehydrogenase" /protein_id="BAB06868.1" /db_xref="GI:10175771" | |
| CDS | /translation="MKVKVAINFGFRIGRMVFRKAVENKQDVVAINASYPAAETLAHL IKYDSVHPFGQTVEIKNDALVNGKEIKLVNSRDPVLPKDKLEVDIVVEATCKFKT KETAGYHLEAGAKKVIITAPCKNEDTVVGVGNENDYKPEHHVLSNASCTTNCGLAPV AKVDEAFTTQSGMMTTHAYTNDQKNIDPHKDLRRARACGQSIIITPTSGAARAKISK VLPQEGKLTGMALRVPVTPNVSVDLVYQVKEDVTPEKVNHVLRKOAAGSGMKGLIGFT EELVSTDGNGNEHSAIVDGLSTWMDTNTQVKVLAWYDNEWGYSTRVVDLVELVAKE LKKPAKVIAS" /complement(268822..269427) /gene="BH3150" complement(268822..269427) /gene="BH3150" /note="BH3150 unknown conserved protein" /codon_start=1 /transl_table=11 /protein_id="BAB06869.1" /db_xref="GI:10175772" | |
| gene | /translation="MMIGLTGGIAGSKSSVAKMMEELGLPIVDADQVARDVVEPGMPA YEATVAFHGQVNVNDCGTLNARKALGSIVFOQEERRVLEIVHPAVBRQMQOQKEOLI RSGETIVFDLPLLYESNLRYLVEKVLVLYVDEHTQQLRMNRQAKGKDAIHRIRSQ RPLESKDRADATIDNSGTLTDATRKQIDILKRWKVIPEQO" /complement(269453..270094) /gene="BH3151" complement(269453..270094) /gene="BH3151" /note="BH3151 unknown conserved protein in B. subtilis" /codon_start=1 /transl_table=11 /protein_id="BAB06870.1" /db_xref="GI:10175773" | |
| CDS | /translation="MIHLVSLALAFVSLDSFCVGLTYGLRKKMKIPFKSLFFTAGCS AVSLIAMGFGMTIQYLSFAIAESLGGIILIGLWALYQAFPAKNRGHKTKDDHVI LNFEIKMGVVIRILRKPMWADNSGAITGREAVILGIALSLDAFGAGIGAAALIGPH PLYMAVSAYMSALFVSLGKMSGSRFADSYWIKRFSFIPGILLVLIGVFNLM" /complement(270162..270986) /gene="mutM" complement(270162..270986) /gene="mutM" /EC_number="3.2.2.23" /note="BH3152" /codon_start=1 /transl_table=11 /product="formamidopyrimidine-DNA glycosidase" /protein_id="BAB06871.1" /db_xref="GI:10175774" | |
| gene | /translation="MPELPEVTVRRRTLAEVLIGKTIQVDVGVGWAKMIKRPDDVDQFK WLLKGQTRISGMGRGKFLFLHDDYTLVSHLRMEGRYGLYQONESVAKHTHVRFVGD GTLELYQDVRKFGTMHLFQSRQOMEPLAKLGVPEPSDOFSAKLLTERLSKTSRKIK SALLDQGIIVGLGNIYDEALFRARIHPERLAKDVTVAEVKILHQAILNLTIEAVNLG GSSIKSTVNGOGENGMEFQQRLDYVGRKGETCRQCGTPTITVVGGRGTHFCVCQCK" /complement(271009..273639) /gene="poIA" complement(271009..273639) /gene="poIA" /EC_number="2.7.7.7" /note="BH3153" /codon_start=1 /transl_table=11 | |
| gene | /translation="MTKFERVLAVITVTVLLVMAGLGLVIGQIFKNVYLENLDRUK KETYLASMVNEAVEALFNEVQTLTEESQKLDARVTIILADGTVVGSAADPAEMNH ADRPETLEPEGIVRYSTVTELLFYAVIQNEANETIGYVRGLGRTAEVNSVRNRL WAILIVSTTAFLVIVSVTVRIANQMRTESATVAVANKLAEGDYQARTSESRSRDE QLNRSINVLAYNLQTLKRQVOKERLETLEIENMGSGLLINTRGDLSLINTKTCHDIF QEDTDLWLHQLYHDVIRKKEIILKIQVDILFEKKRQKVKLPIHLEYRHFDDVHGAPIV RENGKLGIALVFHDITLKKLEQVRKDFVANVSHELKTPVTSIKGTFTLLDGMHD EQLRDQFLHIIWKESERLSLIHDLLELSKTEQNYFQLNMQQTNLFVAVSEVMTLLKG KABEKGIDISLSEBSGSPLEGDPERLKQIAINLVNNAITVTSNGRIDLALKDHGDV | |
| CDS | /transl_table=11 /product="two-component sensor histidine kinase involved in phosphate regulation" /protein_id="BAB06875.1" /db_xref="GI:10175778" /translation="MTKFERVLAVITVTVLLVMAGLGLVIGQIFKNVYLENLDRUK KETYLASMVNEAVEALFNEVQTLTEESQKLDARVTIILADGTVVGSAADPAEMNH ADRPETLEPEGIVRYSTVTELLFYAVIQNEANETIGYVRGLGRTAEVNSVRNRL WAILIVSTTAFLVIVSVTVRIANQMRTESATVAVANKLAEGDYQARTSESRSRDE QLNRSINVLAYNLQTLKRQVOKERLETLEIENMGSGLLINTRGDLSLINTKTCHDIF QEDTDLWLHQLYHDVIRKKEIILKIQVDILFEKKRQKVKLPIHLEYRHFDDVHGAPIV RENGKLGIALVFHDITLKKLEQVRKDFVANVSHELKTPVTSIKGTFTLLDGMHD EQLRDQFLHIIWKESERLSLIHDLLELSKTEQNYFQLNMQQTNLFVAVSEVMTLLKG KABEKGIDISLSEBSGSPLEGDPERLKQIAINLVNNAITVTSNGRIDLALKDHGDV | |

| | | |
|------|--|--|
| gene | /product="DNA polymerase I" /protein_id="BAB06872.1" /db_xref="GI:10175775" /translation="MKKLVLDGNSIAYRAFFALPLLNDKGVTYNAVYGTMMLLKV LEDEKPTHMVAFADGAKTTFRHDTYDAYKSGRQKTPPSESEQFLIRELLDAFNIARV EVENFEADDLIGTLAKASQNFVKITGDKDLUQIVSDHVTVALTRKGITQIDVID EKGEERYGTPEQIIDMKGLMGSSDNIIPGVPGIGETALKLLKEYSGSVKLVLESID EISGKMKERLEHRDQALSKLATIYLEAPVTLADASFTFGYDKRKEVVKLFKEILD FNSLDKVDADGEGIEKEELQSLNVTLEVRKEAHLDAALVVEVDENYHOGDILG LAJAHNNGCYVPTDEALRSFALQMLDKBERKEKLPDAKRAVVALGKNIELAGSTF DLQWASYLLDDPSTSFHEVSDVTKHHRKGHVHEDENYCGKAKRRIPODEKELGEHLAK ADAIPOKEHVEEKLEKENEQGLFENLEMPUSVILGRMETTGVKVDVEQLQSGMEDLN ARLEEQTTHELAGVSFNINSPKOLGEILFEKVLQIPVKTKTKGTSYTSADVLEKLAG QIETIDHILRYQLGKLNSTYIDGLLKVVHKGTKVHTRFNQALQTQGRLSSTDPNQL NIPTRLREGKIRKAFVFPSEKDFILAAQDSQIELRVLAHFAEDEKLMKEAFREDMDVH TKTAMDVPHVSOEEVTSNMRSAKANFGIVGGISDYGLSONLGITRKEAATFTERV ASYPKVXEYLDHVIDEAREKGYYVTLLKRRRYLPDLTSRNFNQSPFAERTAMNTPIQG TAADIIKOAMVQMASRMDDEEDLQSLRLQLVHDELIFEVPREEVAOMKELVTSVMEQAV ELRVPLKDVACGDTWYDAK" complement(273957..274889) /gene="hfIC" complement(273957..274889) /gene="hfIC" /note="BH3154" /codon_start=1 /transl_table=11 /product="protease specific for phage lambda cII repressor" /protein_id="BAB06873.1" /db_xref="GI:10175776" | |
| | /translation="MSDDKIINFDEKRPVELRKYVQLGVSVAVLIGIVGILSNLFI VEQEGYKVVROFGVWRVRESPPGLKFKIPTIQSVSTLPKYOMIYDIPPAEINTRDKKR MMADHYALRIEDPLRMISNVSGLQGAELIGEOLFSARAELGQLEPGLINEEENS RGDNQVKERVNSLSERQDLGIVLLDVRMARTDLPKENEERAYRMIRESERISADQY LSQGDANRIRARTDQEVTEILAKADAEELIIGAGEAAAEIYNESFGNDPPEFYD YRTLLSYEKTIGDQTVIVLPADSPYARILMGYTD" complement(274882..275841) /gene="hfIK" complement(274882..275841) /gene="hfIK" /note="BH3155" /codon_start=1 /transl_table=11 /product="protease specific for phage lambda cII repressor" /protein_id="BAB06874.1" /db_xref="GI:10175777" | |
| gene | /translation="MTIRQLVVGFFSLIGAAILGLFLVTGWYIVDETEQAALITFGKV EETIDEPGLAKFMPWPQKVLEILPRGFNLQVGYKEDEGEVVEFTDEAKMITGDENIV FADLAVQWRITDPEOYLYSTEDPKELLYNATSSALRSVIGSAYDEALTDSPRTIEAD IFESLVELMDLYQIGISIDSVKLDQVELPTEEVRAFTDVTDAEREELTKINEANRYR NQETNEVEGEKDALISRAEGORADRIETARGDVARFNALYEEYLVNPDVTHQRLVLET LESILPDTEIYIMDSNNDTINYLPIRLEROQQAPVEEGTNN" complement(276268..278037) /gene="phoR" complement(276268..278037) /gene="phoR" /note="BH3156" /codon_start=1 /transl_table=11 /product="two-component sensor histidine kinase involved in phosphate regulation" /protein_id="BAB06875.1" /db_xref="GI:10175778" | |
| | /translation="MTKFERVLAVITVTVLLVMAGLGLVIGQIFKNVYLENLDRUK KETYLASMVNEAVEALFNEVQTLTEESQKLDARVTIILADGTVVGSAADPAEMNH ADRPETLEPEGIVRYSTVTELLFYAVIQNEANETIGYVRGLGRTAEVNSVRNRL WAILIVSTTAFLVIVSVTVRIANQMRTESATVAVANKLAEGDYQARTSESRSRDE QLNRSINVLAYNLQTLKRQVOKERLETLEIENMGSGLLINTRGDLSLINTKTCHDIF QEDTDLWLHQLYHDVIRKKEIILKIQVDILFEKKRQKVKLPIHLEYRHFDDVHGAPIV RENGKLGIALVFHDITLKKLEQVRKDFVANVSHELKTPVTSIKGTFTLLDGMHD EQLRDQFLHIIWKESERLSLIHDLLELSKTEQNYFQLNMQQTNLFVAVSEVMTLLKG KABEKGIDISLSEBSGSPLEGDPERLKQIAINLVNNAITVTSNGRIDLALKDHGDV | |


```

subunit"
/protein_id="BAB06884.1"
/db_xref="GI:10175787"
/translation="MSTELDFEKPTRDLKGKIEELRTFTTEKEIDLSDIEKLEKELH
ALEENLYGNLKPQWRQVQIARHGERPTTLDYIEQLSDFLEMHGDRLYGDRDEAIVAGIA
KTKQPVTVIGHQRQDKTENIRNRPFGPEGYKRALMKQAKFHRPVICFIDTK
GAYPKASERGOSAKIAARNLLEMAGLKVPICVIVIGEGSGGALAGVGRDTHMLEN
STYSPISPEGAALLMKASQAQAAETMKITAPDLKELQIIDDIPEVRGGAHNRVA
EOAEAIQDVLEKSLKOLSSLTTEELLNKKRYEKYKIGEEFHVNDALISVN"
gene
complement(287757..288605)
/genes="accD"
CDS
complement(287757..288605)
/genes="accD"
/EC_number="6.4.1.2"
/notes="BH3166"
/codon_start=1
/transl_table=1
/product="acetyl-CoA carboxylase transferase beta subunit"
/protein_id="BAB06885.1"
/db_xref="GI:10175788"
/translation="MLKDFSKRYATIPSERAKHDIPEGIMTKPCSCRTIMYTKEL
KNLYVCDSCGFHHRMNAFDRIESLLDPCTFVELDKGMTEDPLSPPTYREKVEADRK
KTNLAETVIGSGTNGFPLVIGVMDARPMGSMGVSVCGETRALERAIEHNKQPELL
FNASGARMQEGVLSLMQAKTSALERLDRVGGLFISVWHTPTTGGVSASFASGDY
NPAEPGALGFAGRIIEQTIREELPNDFQTAEFLLKHGQLDRVIPRSEMKOTLTKIL
DIHVS"
gene
complement(288671..289324)
/genes="BH3167"
CDS
complement(288671..289324)
/genes="BH3167"
/codon_start=1
/transl_table=1
/product="transcriptional regulator (GntR family)"
/protein_id="BAB06886.1"
/db_xref="GI:10175789"
/translation="MSAGDKVYLEVIQINQIIRADQLSAGDKLPSERELSRHLHVR
SSVRALRALLEDLDIRREGCTFIRPANSYPLVSFLSKDQOALKDLGETRRI
VIEIARLACRRITETIQOLDLEILSEIEKHSEDDFRKHDFRHERFIVAACQNRL
YNMASLVEYHOTAMTYEVVQRRORKAIVQEHRTIVEALKKKDEQATLTMALHLQH"
gene
complement(289365..290597)
/genes="BH3168"
CDS
complement(289365..290597)
/genes="BH3168"
/EC_number="1.1.1.38"
/codon_start=1
/transl_table=1
/product="malic enzyme (malate dehydrogenase)"
/protein_id="BAB06887.1"
/db_xref="GI:10175790"
/translation="MATTREEAHLIHRENOGKLESKVPVVRNARDLSLAYSPGVAEP
CKEIFEDVTVYETMKGNMVAVSNGTAVLGLNIGPEASLPVMBGKAVLFKSPAGV
DAFPICLRTTVDQIVETVKLLEPTFGVNLEDIAAPSCFEIERLKKETNIPVPHDD
OHGTAIVTLAGLINALKLTGKSMELKVVANGAGAGIAIILKQRMGVRDIIMCDSK
GAIYDGRPYGMNDMKAEVAKWTNRKIDGGLGVLGADVFICGVSVAGALTREMIENM
NDPPIIFAMANVPVEIMPEEAKAAGAKVIGTGRSDFPNQNVNVLAPFPPIFRGALDARA
THINEMKVAAYHAITADLVAEEDLPDYVIPDFPRVAPAVAAAVAKTAMETGVARR
KVDPAEVEAKTKKLAIE"
gene
complement(290740..294090)
/genes="dnaE"
CDS
complement(290740..294090)
/genes="dnaE"
/EC_number="2.7.7.7"
/notes="BH3169"
/codon_start=1
/transl_table=1
/product="DNA polymerase III alpha subunit"
/protein_id="BAB06888.1"
/db_xref="GI:10175791"
/translation="MESVHIHVHSEYTLTSSMCRIPALVEKAKAAOFALALTDRHVM
YGAVPYKACKESGLKPIIGMETTVRFADRESDLLLYARSNKGVEHLLKLTILQCR
EKEKYVTHEELLAYKDDLIIYVIPYSGGVFRACLEALERGERFTQFLKDLHGGESV
YLEIQTGRSKVEPINQLAAKLAIPLVCSQHIQVLEREDLDANKVIOAIRREGVLYVEEL
RIEQEEDVCFFTLTMEQTFRAYPEALQNTKKLADRCHVELTIGKPRLPKFQTPNEQS

```

```

AEDYLKLCCECAKERYGEEWTKEAQRIGEELAVIERMGFSDYFLIWMDFMRFAREA
SIVTVGPGSVAGSLVAYVLFTIDVDPLHYDLFLFERLNPERSLSDIDIDFDPHRRR
EVIVYVKKYGANRVAQILFTFGLAARASVRDVKALAIIPNVIEKISEISGRPGMT
LVKAYNENRLQOLHASDEAOKVMKLARKVEGLPRTTSTHAAGVVISQOPLTEVIAL
QHGQGEVPLITQGTMETVEDVGLIKMDFLGLNLITLLEMVVKVRREYVGHSLNVKQLPL
NOAKTETALLAKGETTGVFOLESGMRKVLRELKPNTFADIVANALYRPGMEFIPDY
IQCKEGTKITTYLHPDLEPILSTGYIVYEQIIMAAKMGFSGEADLLRRAISK
KKEALRQOEFAAVFTGAVRQGYEQETAKIYELIVRFANYGFNKSHAVAYSMLAYOLA
YLKAHYPSGFYAALASTIWNQPEKLERLLQEMKQOQIRVLPPLSKSDIHFSEEGEV
RPLPLPRLYVSVAIRELJKAREEAPVRSFLDLCSRVDGRIVTSRVWESLJKAGALDE
LGERATLLANIEEAFQAEQVEPOENTGGLFQLSVEEPEYIKVEPLTDLKLAYEKE
AVGFYLSGHPLLAYTESLQYDRLTYLEGTFERRFVAKLMIHRIRRTKRGVEVMGFL
TMSDETGEWEAVVFPAPWAMYEWLKEGELYFVEGMDRGRSEELQLLDVLDKVLPLKHM
LKKEKEKFLFKITADYKEVERLNDIRRLILQHVHGPTPVVVMYKEQRKTKQLPEKYHV
LSLSFTLLHELOLVGKEHVVVSTYDEES"
BASE COUNT 79260 a 71009 c 57462 g 86519 t
ORIGIN
AP001517 Length: 294250 February 15, 2002 13:42 Type: N Check: 8703
Found using 'seq4' (pappu403.key)
...
350 TTATTAAGTGAGCAATACGGCGTCCCTATGGAAGCTATTATAACGATATTAAACGACGTTCT
400 407
410 TCCAATACTATTATTCGTGGAGAACAGCTAACAAATCCCTGCACGGGTA
...
6574 GTTCATTTCATACTACATAAATGGGTGATTGAATCACGTCGATGATAAAAACGTTCCGT
6624 6631
6634 TGTCACAGCTGCTTCATACCTACACGCGTATCCGCTTATTAATCGGGCC
...
21118 TCATCTAACGGTAACCCCTCTTCTGTGATGATGAATAGCAGATGATTTCTCTCAACGTCGGA
21168 21175
21178 TGCTGGGATAACAGACGAATGAGTTCTAGTCTCTCCGTAGCCTGTTGCA
...
31246 TTACGATGAAATGGAAGTTGAGACAGCACGCTGATGAATCTTTGCAATCGTCGAACGTTTG
31296 31303
31306 GGCACGCTGGCTCCAACAGCCTTCATATCAACGGCTGCTGCTCACTAAC
...
80129 GCATCGGGGATCACCCCTCTTTTCATCAACACTTCGTTTCATAAACTGGAACGCTTCGTT
80179 80186
80189 CGAACATGGATCGTACCATAACAATCCCTTTTTCGCTCGTATCGTTGTA
...
84441 GATCAGACCGCCTAGTCAGTCCTCTCGTTTGTATCCGCTTTTCATTTATGAAACGTTTCGCA
84491 84498
84501 CCATTGTCAATAAACTACTGATGACACTCTCACTTAATCTACTTCTCTCC
...
100168 CGTTTCATCTGTGATGAAAAATGTCTTTCTGGCGCCCTCGGTCTCATGTTGGAACGTTTCGCA

```

100218 100225
100228 AGCCTTTTAAACCTTTATCCGAAGAAGATGTGGCACTTGGCTATCGC
...
100549 CGTGCAGGTGAGTTCGTATTGTGACCTGCTAGTGACGTGACGAGGAACGTTCCGAG
100599 100606
100609 GAAATTTACGGTAGAAAAAATTCGCGAAGATGACCTAGTAGGAGGTGACG
...
118176 AACCATAGACGATGGTGGGAATTCGCCCAACACTTCGAGCATTTGGCTTCAACGTTCCGCC
118226 118233
118236 GCAAGCATCAGACGCATATTCTACTTAAAAAAGATCGCTGTCTATTAAAC
...
119543 AGAAATTTACCAATCCAGCACTTGTTCACGCGCGGCTTACCTCTACTGAAACGTTCCGT
119593 119600
119603 TGCTCGCTTATCATATACTCTTCAGCAATGGCTGCCATTAAACGGATAA
...
164326 CCCCCTGGCATGAGCGGCTTCGCTCGATTAAACGCTGCCCTTGATTTTTCGACGTTT
164376 164383
164386 CATAGCATATCCCCCTCACTTCTCTTCATTAAACCAACGGGCTACAT
...
181749 TCTTGTGCCACTTCGTTTGGCCACTTCTCGCCACATCGGACTCGATTCAAGAACGTTCCGC
181799 181806
181809 TTGTCAACCGATGTGACATGTTTATTTCGAACCATGGCTAGCTCAAG
...
221934 TGCTTCCGTTAAGCCGTATCTTCCACAGGCGACCTCCTGTAACTGTTTCGAACTGTTG
221984 221991
221994 TTGAACCTCTGCTGGGAGAGAGAAGACCCACTGATACAGGCCTCAAT
...
222206 AGTTCATCGACGTTGTTCATGCCATACACATGGAAGATGGCAATACAGCTAACGTTCCGT
222256 222263
222266 CTTGCCCATCGCGCCATTGTGTATCCATTTCGGCACTGAATCGTAT
...
229541 TCGCGGCTAAAATTTTTTCTTCAGACTTCTTCCCAATCCGGCTAATTCGCGAACGTTTC
229591 229598
229601 CTTGCTCACATGCTCCTCTTAAATGTAGCAAGATCAATCACCCTAGCT
...
262990 GCGACAGTACGTATCGCAAAAGTTTGTATAGTTTCAGGTTCCGCTCCGATAAACGACGTTCT
263040 263047

263050 TATTGGCTCGGGTGCCCATTTTAAACATGATTTGATTGGCGGCTGGCGC
...
278558 TACATCGAGTCCATCCATTTCCAGGTAACATCAGGTCTTAAATATAATAGATCGAACGTTTG
178608 278615
278618 CGTCTAGCTAGTTGTAGCCAGAGCTCCATCCATTCCTGTCACGAC
...
10 matches found in sequence:
ap002554 ; TOIG of: ap002554 check: 6396 from: 1 to: 327773
(from "immun_ge.seq")
TOIG of: ap002554 check: 6396 from: 1 to: 327773
LOCUS AP002554 327773 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION AP002554 BA000007
VERSION AP002554.1 GI:13360491
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
REFERENCE 2 (sites)
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MGI655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
REFERENCE 3 (sites)
AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
REFERENCE 4 (sites)
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
REFERENCE 5 (bases 1 to 327773)
AUTHORS Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome

Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:kengen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)

COMMENT
FEATURES

Location/Qualifiers

1. .32773

/organism="Escherichia coli O157:H7"

/strain="O157:H7"

/sub_strain="RIND 0509952"

/db_xref="taxon:83334"

150. .2258

/gene="Ecs1032"

150. .2258

/gene="Ecs1032"

/note="similar to VCBY_ECOLI gi|1787181 percent identity

99 in 702 aa (Conserved in E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="putative oxidoreductase"

/protein_id="BAB34455.1"

/db_xref="GI:13360492"

/translation="MNSLFASTARGLEELKTELENLGAVECVQGVGHFKGDTRLV
YQSLMRSASRIIMPLGCEKVYSDLDLYLGVOAINTNEMFGATFAVHFSGLNDTI
RNSQYAKWKVDAIVDAFTKKNLPRPNVDAPDIIRVNMVHLKRTASIALDLSDGDLH
LRCYDRAGTAPKETLAAAVMRSGWPGCTPLDDPMCGSTLLIEAAMLATAPGL
HGRMGFSWAQDEALINQEVKAEATARKGLAESHYFGSDSDARVLOARTWAR
LAGIGELTFEYKDVQALNPLKEPYGVVLSNPPYGERLDSPELIALHSLGLRIMK
NOFGGNLFLSFDLLSCQLRADQYKAKNGKPLDCVOKNKHVAEFTSPKAMVA
EDYANKRLNKKFEKWARQEGIECYRLYDADLPEYNVAVDRIADVVVQVEYAPPKTI
DAHKARQLFDIIAATISVLGIAPNKLVLKTRERQKNGYQKLGKGEGLVEYTEYNA
HLWNNLYDLTDGLDRIARRMLGOMSKGDKFLNFSYTGSAITVHAGLGGARSTTT
VMSRTYLEWAEARNLRLNGITGRAHLIQADCLAWLREANEQDFLIFDPPTNSKR
MEDAFVDQDHLAKMDLRLRLLRAGTIMEPNSNKRGRFMDLDGLAKLGLKAQITOKT
LSQDFARNQIHNCMLITAA"

2270. .4177

/gene="Ecs1033"

2270. .4177

/gene="Ecs1033"

/note="similar to UUP_ECOLI gi|1787182 percent identity 99

in 635 aa (Conserved in E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="putative ATP-binding component of a transport

system"

/protein_id="BAB34456.1"

/db_xref="GI:13360493"

/translation="MSLSIMHGAWLSFSDAPLLDNAELHIEDNERNVCLVGRGAGKST
LMKILNREOGLDDGRIIYEODLIVARLOQPPRNVEGVSVDFAEGTEEOAEYLKRYH
DISRLVWNPDSERKNELAKVQEOLDHNLWOLENRINEVLAOLGLDPNVALSSLGG
WLRKAALGRALVSNPRVLLDDETHLDTIDWLEGEFLKTFNGTIIIFSHDRSFIRN
MATRIVDLGRKLVTPGNTDQYLLEKEALRVEELQNAEFDRKLAQEEVWIROGKA
RTTRNGSRVRLAKMRREGERREVTAKMQVEEASRSKIVFEMEDVCYQVDGKL
VKDFSAQVLGDKRIALIGPNCCKTTLKMLQGLQADSGRIHVGTKLEVAYVDQHRA
ELRDLKPSNLLIDDEPTNDLDTLELLELIDSYOGTVLLYSHDRQFVDNVTTCW
IFEGGKIGRYGVGYHDARQEQYVALKOPAYKKNEEPAAPKAEVTKRSSKLSYKL
QRELEQLQLELLEAKLEALQTVADASFSPQHEQTQKVLADMAAAEOELEQAFER
WEYLEALKNGG"

4307. .5560

/gene="Ecs1034"

4307. .5560

/gene="Ecs1034"

/note="similar to POIA_ECOLI gi|1787183 percent identity

99 in 417 aa (Conserved in E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="paraquat-inducible protein A"

/protein_id="BAB34457.1"

/db_xref="GI:13360494"

gene

CDS

/translation="MCEHHHAHKLSCQCDMLVALPRLRLEHQAACAPRCGTTLTVMV
DAPQRTAYFAALAEMLLLSNLFPFNMNVAGVTSEITLLEIPGVLFSEDSVASLGTF
FLFVOLVPAFCILITILLVNRLEPVRLEQOLARVLQKLTWGMAREIFLAGVLVSEFV
KLWAYSIGVSSFLPRCLFCVQLQRAFCQVDRWLWDDIAPMPELRQPLKPGVTGIR
QGLRSCCCTAILPADPFCVPCRGTRGIVRRNSLQWTLALLVTSIMLYLPA NILPLIM
VTDLLGSKMPTSLIAGVILLWSEGSYPAAVIFLASIMVPTLKMIAIAWLCWDKAGHG
KRDSERHLLIYEVFEVGRWSMIDVFVIAVLSALVRMGGLMSIYPANGALMFALVVM
TFMSAMTFDPLRSLWDRQPESEHEES"

5565. .7205

/gene="Ecs1035"

5565. .7205

/gene="Ecs1035"

/note="similar to PQIB_ECOLI gi|1787184 percent identity

99 in 546 aa (Conserved in E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="paraquat-inducible protein B"

/protein_id="BAB34458.1"

/db_xref="GI:13360495"

/translation="MESNNGEAKIQVKMSPVPIPTVITALIGAWVLFYHSHQGP
VTLITANAEGIEGKTTIKSRSDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
VFVYVQPIRGREGISGLGTLISGVYTELOPGAGKSKMDKYLDDSPPLAPDAKGI
VLDKSKAGLSPGDPVLFGRIVGSVETSTFDQKRNISITOLFINAPYDLRVTNSVRF
WKDSGIADVTLSAGRMVEMGSLTLLSGGVSFDPGELDGQVPAPKTAFLVYDDQKS
IQDSLTYDHIYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVLPFFAFPMRQTFNDYR
IPVLIIRIEPLKMQLENADVVEHLGELLKRLGSLGTNLVTGALYVDLDFYPMT
PATIGREFNGYIOIIPVSGLAQIOORLMEALDKINKLPLNPMIQATSTLSORT
MKNLQTLDSMNKILASQSQOQLPTDMOSTRELNSMQGFPQGSAAYNKMWADMQL
DQVRELQPLVTLNEKSNALVFEAKDKDPEPKRAQ"

7217. .7765

/gene="Ecs1036"

7217. .7765

/gene="Ecs1036"

/note="similar to YMBA_ECOLI gi|1787185 percent identity

98 in 182 aa (Conserved in E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="hypothetical protein"

/protein_id="BAB34459.1"

/db_xref="GI:13360496"

/translation="MTIAALWLAGSSGEINKNYQLPVVOSGTQSTASOGNRLWVE
QVAPDYLAGNVYVOTSDYKYVIANNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
PLGSAOITLVNTVTEFNGRYDGKIVSGVSWLLNHQGLIKRPRFLEGVQTDGVDENV
KVLAGVWSQEAASIAQBIKRLP"

8021. .8188

/gene="Ecs1037"

8021. .8188

/gene="Ecs1037"

/note="identical to RMF_ECOLI gi|1787186 (Conserved in

E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="ribosome modulation

/protein_id="BAB34460.1"

/db_xref="GI:13360497"

/translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGW
REAMADRVNMA"

complement(8258. .8776)

/gene="Ecs1038"

complement(8258. .8776)

/gene="Ecs1038"

/note="similar to FABA_ECOLI gi|1787187 percent identity

100 in 172 aa (Conserved in E.coli K-12)"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="beta-hydroxydecanoyl thioester dehydrase"

/protein_id="BAB34461.1"

/db_xref="GI:13360498"

/translation="MVDKRESYTKEDLLASGRGELFGAKGPQLPAPNMLMMDRVVKMT

BTGNFDFKGYVEAEIDINPDILWFFGCHFIGDPVMPGCLGLDAMMOLVGFYLGWLGEG
 KRALGVCEVFTGQVLPFTAKVTVRIHFKRIVNRRLINGLADGEVLVDGRLIYTASD
 LKVLGFDQTSAP"
 /complement(8845. .10605)
 /gene="Ecs1039"
 /complement(8845. .10605)
 /gene="Ecs1039"
 /note="similar to B0955_ECOLI gi|1787188 percent identity
 99 in 586 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative ATP-dependent protease"
 /protein_id="BAB34462.1"
 /db_xref="GI:13360499"
 /translation="MTITKLAWRLDLPDTSYQEIFAQPHLIDENDPLFSDTPRLQF
 ALEQLHTRASSFMKAPESEYLNLIANAARTLQSDAGQLVGGHYEVSGHRIRLR
 HAVSADONFATLTQVAAADWVAEQLCFQFNGDITLQPLVHQANGGILLIISLRT
 LLAQPLLMWRLKNTVNRERFDWAFDSRPLPVSPVSMPLKLVLLYGERESLADFOE
 MEPELSQALYSEFEDTLQIVDAESVQWCWRWFTTARHNLHPAGGADAMPVLIREAA
 RYTGQETPLSPQWILLROCKEVAASLDCDGTDFSGEQLNMLQOQREWREGFLAERMDE
 ILQOILIEETGERIGQINALSVIEFFGHPRAFGEPSRISCVVHIGDGEFTDIERKAE
 LGNIHAKMMIMQAFMLSEQLQEQIIPFSASLTFEQSYSEVDGDSASMAELCALISA
 LADVPVQSAITGTSVDFGQRAOPVGLNEKIEGFEFFAICQORELTGKQVLIPTANVR
 HLSHSELVKAVERGKFTIWAVDVTDALPILLNLVWDGEGQTTLMQTIQERIAQASQ
 QEGRRFPWPLRWLWFIPN"
 10791. .11243
 /gene="Ecs1040"
 10791. .11243
 /gene="Ecs1040"
 /note="similar to YCBG_ECOLI gi|1787190 percent identity
 100 in 150 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative dehydrogenase"
 /protein_id="BAB34463.1"
 /db_xref="GI:13360500"
 /translation="MKYQLENLSGKWKYLYKKHREGELITRYIEASAAQEAVDVL
 LSENEPVLVNGWIDKHMPVNLNRMKQTTIARRKKRHFNAEHQHTRKKSIDLEFIVWQ
 RLAGLAQRKGTSETIVOLIEDAENKERYANKRSSLKQDLQALLGRE"
 complement(11319. .12359)
 /gene="Ecs1041"
 complement(11319. .12359)
 /gene="Ecs1041"
 /note="similar to OMPA_ECOLI gi|1787191 percent identity
 100 in 346 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="outer membrane protein 3a"
 /protein_id="BAB34464.1"
 /db_xref="GI:13360501"
 /translation="MRKTAIAIAVALAGFATVAQAAPKDNWTYTGAKLGWSQYHDTGF
 INNNGPTHENOLGAGAGGYOVNYPVGFEMGYDMLGRMPYKGSVENCAQGVOLTA
 KLYGPIITDLDIYTRLGMVMWRADTKSNYVGKNDTGVSPVAGGVEYAITPIAIRL
 EYQWTNNIGDAHFTICRPDNGMLSLGYSVRFQCGGEAAPVAPAPAPEVQTHKFTLK
 SDVLFNFKATLAPQEAADQLYSQSLNDPDKGSVVVLGYTDRTIGSDAYNOGLSER
 KAGSVVDYLISGKIPADKISARGMGESNPVTGNTCDNVKQRAALIDCLAPDRREIEV
 GKIKDVVTQPA"
 complement(12716. .13225)
 /gene="Ecs1042"
 complement(12716. .13225)
 /gene="Ecs1042"
 /note="similar to SULA_ECOLI gi|1787192 percent identity
 100 in 169 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="suppressor of lon"
 /protein_id="BAB34465.1"
 /db_xref="GI:13360502"
 /translation="MYTSGYAHRSSFSFSSAASKIARVSTENTTAGLISEVYVREDQPM

MTQLLLLLLOQLGQOSRWQLWLTPQOKLSREWVOASGLPLTKVWQISQLSPCHTVES
 MYRLARTGNYSVVIGWLADDTTEEBHAELVDAANEENAMGFTIMRPVSASSHATROLSG
 LKIHSNLYH"
 13444. .14073
 /gene="Ecs1043"
 13444. .14073
 /gene="Ecs1043"
 /note="similar to B0959_ECOLI gi|1787193 percent identity
 99 in 209 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34466.1"
 /db_xref="GI:13360503"
 /translation="MKSLSYKRIYKSOBYLATLGTIEYRSLSFGSYSLTVDVTFAMVVS
 DGELYLRACEQSAOCYKVPVWLTYKCGRSVTLNYYRVDESRLWNLQKLVLRLSKYS
 LDAALKKSTRNTRERLKDLPNMSFHLNLTGEGVIGKDVRLALRILGAKMCLWLRQQN
 SLVTEKILFLMEGALIGIHEAALPVARRQELAEWADSLTPKQEPAAELE"
 complement(14036. .16198)
 /gene="Ecs1044"
 complement(14036. .16198)
 /gene="Ecs1044"
 /note="similar to B0960_ECOLI gi|1787194 percent identity
 99 in 720 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34467.1"
 /db_xref="GI:13360504"
 /translation="MAFMLSPLLRKYTNWSAWLYARIFALCGTTAPFPMWLGDKVLT
 IPLTLGVAALATDLDRLAGLRNLIIITLFCFIASASVELLPWPLVFAIGLTLST
 SGFTLLGGLGORVATIAFGALLIAYITMLGTSLVHEWVQOQPMYLLAGAWVNVLTILG
 HLLFPVPRLODNLARCYEOLARYLELSRMPDPIEDESQAPLYDLALANGLLMATLN
 QTKLSLITRLGRDQGRTRTLHYFYVAQDIHERASSSHIQOYTLREHFRHSVDLFR
 FQRLMSQGCQACQSLSRCILLRQPYQHPHFERAFTHIDALERMONGAPADLLKTL
 GFLNNLRAIDAQALTIIEQAQALPHNDENELADSDPHGLSDIWLRLSRHFTPESA
 LEFHAVMSVLVCFYAIQIITGMHGHWIILLTSLFVCQPNYNATRLRLKRIITGLV
 GIAIGIPLWFVPSLEGQLVLLVITVYLFRAFNVQVAHATMETITLLVLLCFNLIGEG
 FEVALPVIDTLICGATAAAVASVWPDQOFNRLPMLERATEANCYLDAILQYHQ
 GRONRLAYRIARRDAHNRDAELASVSNMSEPNVTQIIRAAFRLLCLNHNHTFYSIS
 ALGAREQLTNPELLAFDVAVCYIVDDALHQHQPADEERNEALASLKQRMQQOLEPRAD
 SKEPLVVQQVGLLIALLPEIGRLQRTQVFPQETPVSA"
 complement(16208. .16654)
 /gene="Ecs1045"
 complement(16208. .16654)
 /gene="Ecs1045"
 /note="similar to YCCF_ECOLI gi|1787195 percent identity
 99 in 148 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34468.1"
 /db_xref="GI:13360505"
 /translation="MRTVLNINFLVGGFATTILGWLLATLVLSIALIFTPLTRSCWEI
 TKLSLPYGNIAIHVDLENPAGKNVLLNTGCTVLTNIFLWIFFQGWLLCLMHIATGIAQC
 ISIIGIPVGIANEKIAAIALWPGVRRVSVETAQAAREANARRFE"
 16777. .18831
 /gene="Ecs1046"
 16777. .18831
 /gene="Ecs1046"
 /note="similar to HELD_ECOLI gi|1787196 percent identity
 99 in 684 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="DNA helicase IV"
 /protein_id="BAB34469.1"
 /db_xref="GI:13360506"
 /translation="MELKATTLGKRLAHPYDRAVILNAGIKVSGDRHEHVLIPNQLL
 ATHCKRGLVWGELEFVLDPDEKVVRLHGTWGETQRTFYHLLDAHWRWSEMEIASGV

LRQOLDLIATRTGENKWLTRQETSGVQQQIQRQALSALPLPVRNLEEFDMCREAWRKQ
AWLKDESARLQHOQAYTEAMLTETAYDFFRQVESPINPAQARAVVNGHSLVLVLAGA
GSKTSLVVARAGWLLARGEASPEQILLAFGRKAAEEMDERIRERLHTEDITARTFH
ALALHIIOGSKKVPITVSKLENDTAARHELFTIAERKQCKSEKKAQAKGROWITEEMQ
WSVPEGNFQDEKLORLASLRDNRVSLMRHGGCAQAEIMTASAPEIROLFSCRILM
APLLKAKKALKANAVDESLIHQAIIVILEKGRFISPKWHIILVDEFQDISPQRAALL
AALUKNSQTTLPFAGDDWQAIIRFSGAQMSLTAFHENFGEGERCDDLTITRFSNRI
GEVANRFOONPGQLKPLNSLTNGDKKAVTLDESDQDALLDKLSYAKPEERILIL
ARYHHRPASLEKATRPKLIQIDEMTIHASKGQADYVILVGLQSGDGFPAARES
IMEEALLPPVEDPDAEERRLMYVALTRARHRVWALFNKENPSPFVEILKNLDVPVAR
Kp"
complement(18863..19330)
/gene="Ecs1047"
complement(18863..19330)
/gene="Ecs1047"
/note="similar to MGSA_ECOLI g111787197 percent identity
100 in 155 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="methylglyoxal synthase"
/protein_id="BAB34470.1"
/db_xref="GI:13360507"
/translation="WTIMELTTRTLPARKHIALVAHDHCKQMLSWVERHQPLLEQHV
LYATGTTGNLISRATGMNVNAMLSPMGDDQQQVGA LISSEKIDVLIFFWDPLNAVPHD
PDVKALLRLATVWNIPTATNATADFIIOSPHENDAVDILIPDYQRYLADRLK"
complement(19417..20079)
/gene="Ecs1048"
complement(19417..20079)
/gene="Ecs1048"
/note="similar to B0964_ECOLI g111787198 percent identity
100 in 220 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34471.1"
/db_xref="GI:13360508"
/translation="MKTGIVITLTIALCLPVPVFATTLRLSTDVDLLVLDGKVVSSLL
RGADSTELDNGPHOLVFRVEKTHLSNREERLYISPLVVSFNQTLINOVNRLPRLE
NREANHFDAAPLELLDGDATPIPVKLDILAITSTAKTIDIEVEVERYNKSAKRASL
PQFATMADDSTLLSGVSELDIAPPQSQVLTEQRLAIWFKLADPQTRNTFLQWAEKQP
SS"
20252..20665
/gene="Ecs1049"
20252..20665
/gene="Ecs1049"
/note="similar to B0965_ECOLI g111787199 percent identity
99 in 141 aa, N terminus is different from b0965
(Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34472.1"
/db_xref="GI:13360509"
/translation="MKETDIAGILTSTHTIALVGSADKPDPRPSRYRVKYLDDGYHVI
PVSPKVAGKTLLGQCGYGLADVPEKVDWVDRNSEAANGVAQEAIGAIAKTLNMQL
GVINEQAARLADAGLVNVDRCPAIEIRLGLAK"
complement(20710..21078)
/gene="Ecs1050"
complement(20710..21078)
/gene="Ecs1050"
/note="similar to YCCV_ECOLI g111787200 percent identity
100 in 122 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34473.1"
/db_xref="GI:13360510"
/translation="MWNFTLISKVKISREVTMTIASFGIGQQVRRHSLGLGYVVVDID
PVYLSLSEPSDELAVNDELRAAPWYHVMWEDDNGLPVHTYLAEAQLSSELQDEHPQEP

SMDELAQTIRKQLQAPRLRN"
complement(21085..22188)
/gene="Ecs1051"
complement(21085..22188)
/gene="Ecs1051"
/note="similar to B0967_ECOLI g111787201 percent identity
99 in 367 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative oxidoreductase"
/protein_id="BAB34474.1"
/db_xref="GI:13360511"
/translation="MEGKASLGETIDIVDHQKWLARGAYSPASQIRARVWTFDPSES
IDIAFTSRLOQAQKWRDMLAQDGLDSYRLIAGESDGLPGITIDRFNGLVLQLLSA
GAEYQRAALISALQTLYPECAIVRSDSDVAVRKEGMELTOGLVTGELPPALLPIEBHG
MKLLVDIOHGHTGYLDDQRDSRLATRVVENKRVNLNCFSYGTGFEAVSALMGCSQV
SVLTISOEALDIARQNVNELNLSKAEFVRDDVFKLLRTYTRDRGEKFPDVTVMOPPKFV
ENKSLMGACRGYKIDNMLAIQLNNEGILLTFSCSGLMTSDLFQKLIADAADAIGRD
VFIEQFROAADHPVIATYPEGLYLKGFACRMV"
22370..22648
/gene="Ecs1052"
22370..22648
/gene="Ecs1052"
/note="similar to B0968_ECOLI g111787203 percent identity
100 in 92 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34475.1"
/db_xref="GI:13360512"
/translation="MSKVCIIAWYGVGVGVFRYTOYEAKRLGLTGAKRLNDGGSV
EVVACGEGOVKLMQWLKSGSPRSARVERVLSHPHPSGELTDFRIR"
complement(22645..23031)
/gene="Ecs1053"
complement(22645..23031)
/gene="Ecs1053"
/note="similar to YCCK_ECOLI g111787204 percent identity
99 in 128 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative sulfite reductase"
/protein_id="BAB34476.1"
/db_xref="GI:13360513"
/translation="MRGFVFCYTAGSLIKVVIIMLIFEKKEIETDEGYLKESQHS
PLAVIAENEGISLSPEHWEVVRFDYLFNFNTSPAIRMLVKAMANKFGEKGNRSY
LYRLFPKPAKQATKIAGLPKPKCI"
complement(23065..23724)
/gene="Ecs1054"
complement(23065..23724)
/gene="Ecs1054"
/note="similar to YCCA_ECOLI g111787205 percent identity
99 in 219 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative carrier/transport protein"
/protein_id="BAB34477.1"
/db_xref="GI:13360514"
/translation="MDRIYSSSHDRTSLLSTHKVLRNTYFLLSLTLAFSAITATSTV
LMLPSQLLITLVGMGLMPLTYKMAKPTGIISAFATGFLGYILGPILNTVLSAGM
GDVIANALGGTALVPFCCSAYVLTRKDSMAGLIVVVLGHWANIFLQLPAL
HLAISAVFLIISGAILFETSNIIHGGETNYIRATVSLYVSLNIFVLSLILGFASR
D"
complement(24132..25151)
/gene="Ecs1055"
complement(24132..25151)
/gene="Ecs1055"
/note="probable integrasesimilar to integrases e.g.
[Bacteriophage HK022] g11385601sp164071VINT_BPDKO
percent identity 27 in 321 aa"

```

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative integrase"
/protein_id="BAB34478.1"
/db_xref="GI:13360515"
/translation="MGRPRKNDNVLPVRVRSNGYSYVWKPEGSTRSIGLGRVRKTS
VAKVQWYEKAKLHMTVAKLHMFMDSPAFTEAPLAPQKDYRQHOKALLMVFQK
VLADNVKTEQVRIFMDKRGLESQTQANHELASLSRVYGGYERGVKNPCKGYRKFS
LKARTVITDEQAAIYAEEAIQPLRIAMEISYLCARLGDVLELKWQDIMGKGIYIQ
NKTGTOKIWEKSPRLRTAIOLARNVSSCTCEYVINTFKGKGIYAKTLNNWNAKRAA
EOKGVPGFCNFHDIAKAGISDYEGSSRDQIIPSGHKTENQVLLYDRKTKITPTLDLP
LVYSK"
25244..25396
/gene="ECs1056"
CDS
25244..25396
/gene="ECs1056"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34479.1"
/db_xref="GI:13360516"
/translation="MGSFLGPITNGMPCSLNWCCPFYRSYFVISLISAGDRVNSYMW
SRAFTA"
complement(25439..27910)
/gene="ECs1057"
complement(25439..27910)
/gene="ECs1057"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34480.1"
/db_xref="GI:13360517"
/translation="MSKIFICAAIPDEQAIKEDSNAVAVATAIAGDERRARRAKFHQF
LEQFPAQCAYKFIVCEDKPGIPRALDSWDTFYMOENRWDEESAFVPEVEDPM
NWFNKLSPVQNAVILVKEDTCENITVDVMSIAQELQEDMATFGHIVEALMKLPEV
NAMYPELKHAIQVWRHKCEPGAKWEIQAEMRIWKRREGERKETGKYSVVLDLTRA
RVNQNTENAAKTGAVTVAIRREYKQIWTKTLDNELACALPGDVGADNIDGTIHWRA
TNEVIDKREDWKRISASMRKQPEALGYDROTIEGLVREPIDIHKDPVALNKYISEY
LTTKVFHEETDQSSDALQPSAAQTAPVETAESDTQNEILVEAPSPVERGGPIYF
VFTDKGKGIKRAKLSGLDKALAAAGTEISKEEYFARKNGTYTGLPQNVDTAEDSEQ
SEPVKTADENVKIMQANISQPDADKLLAASGEFVEGIDSPNDPKWKVGIQSRDAE
DQNPVKNQNEPEAQNSPTQNGPEEQPEPAVQOELEKVCYTAGQGTGGNCPDG
AVMGNATYLETPEENOEAQKNDPEMEGTEHLHKENTGSDQYHASNKTGETANPL
IKVNGHETISSTRLWHLMLDLETWKNPDAPINSIACKFPDPATGEMGPESKTIID
LETAGGVDRDRTIKWIKQSREQAQSSILTDEITLDDALLQFREFIDENSGSEFFVQVWG
NGATFQNVILRRSYERQGIQPCPWRYTNDRDVRTMVALGLVMDPDAKTTIPFEGERHNA
LHDARYQAYVSIAIWQKLLPQSADF"
complement(28004..28195)
/gene="ECs1058"
complement(28004..28195)
/gene="ECs1058"
/notes="unknown, similar to hypothetical proteins e.g. YdfD
[Escherichia coli] gi|140587|sp|P29010|YDFD_ECOLI percent
identity 63 in 63 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34481.1"
/db_xref="GI:13360518"
/translation="MNTTFALVTLYVSGESLELVTGLYGSKMCEMAAAAEQKIPGN
CYPVDKTTHTNNNEIPAGL"
complement(28192..28380)
/gene="ECs1059"
complement(28192..28380)
/gene="ECs1059"
/notes="probable cell division inhibition, similar to dicB
[Escherichia coli] gi|2507009|sp|P09557|DICB_ECOLI percent

```

```

identity 54 in 62 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative division inhibition protein"
/protein_id="BAB34482.1"
/db_xref="GI:13360519"
/translation="MERKLSNASTSELRFETIGVTGDKTFIEDAIKORKLEQDLLNEV
CIFSMRLARLLDLQGYKQ"
complement(28566..28618)
/gene="dicF1"
/notes="inhibitor of FtsZ"
/product="dicF antisense RNA"
complement(28566..28618)
28647..28952
/gene="ECs1060"
28647..28952
/gene="ECs1060"
/notes="unknown, TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34483.1"
/db_xref="GI:13360520"
/translation="MSLKRMVNLTTRCPQRGNKPAILTAAPFRGLPQPEAHGRKEN
DGSKGEGRLRALSCYALTFOGYILSVNCQCRILICVTRTTTRSSVLVSIFNSALNGG"
28954..29139
/gene="ECs1061"
28954..29139
/gene="ECs1061"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34484.1"
/db_xref="GI:13360521"
/translation="MPNKKPCACNALSGFLEKGGYIFNCPHEVPHISKLDSITKTP
NOYSDLLNKLNAARD"
complement(29177..29329)
/gene="ECs1062"
complement(29177..29329)
/gene="ECs1062"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34485.1"
/db_xref="GI:13360522"
/translation="MKGTSADFTYSMAADGEPFPPELKITPGNHPMVTIRALENVW
FNPVQQ"
complement(29326..29715)
/gene="ECs1063"
complement(29326..29715)
/gene="ECs1063"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34486.1"
/db_xref="GI:13360523"
/translation="MEFKDLLEIOEIAAHALHQRLENEVELESATKKYIDNMARNVRD
ATGLYSVSTNNQNTETAKRIASVWGFVEEKYSKKEFWTKTKLQSENCHLLKQS
LISMRRKVIQOTDYRNCNLSNELS"
complement(29727..29855)
/gene="ECs1064"
complement(29727..29855)
/gene="ECs1064"
/notes="unknown, similar to YdfB [Escherichia coli]

```

```
gii140585|sp|P29009|YDFB_ECOLI percent identity 100% in 41
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34487.1"
/db_xref="GI:13360524"
/translation="MDFDTIMEKAYEYFEGLAEGEALSFSEFKQALSSSAKSDN"
complement(29857. .30012)
/gene="ECs1065"
complement(29857. .30012)
/gene="ECs1065"
[notes="unknown, identical to ydfa [Escherichia coli]]
gii140584|sp|P29008|YDFA_ECOLI percent identity 100 in 51
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34488.1"
/db_xref="GI:13360525"
/translation="MDTIDLGNNEslVYGVPNPQDGTFTAMTYTKSKFTKENGARRR
LERNSE"
complement(30070. .30288)
/gene="ECs1066"
complement(30070. .30288)
/gene="ECs1066"
[notes="unknown, TTG start"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34489.1"
/db_xref="GI:13360526"
/translation="MNNFHPFEGFCFKQQLKTVDTKTVFYVCRLLHTPTTHPAPONA
GGYFELPSSGGQVSSQPEAYEHDRVOL"
complement(30290. .30577)
/gene="ECs1067"
complement(30290. .30577)
/gene="ECs1067"
[notes="unknown, similar to hypothetical proteins e.g. yacB
[Plasmid Colib-p9] gii4512441|dbj|BAA75090.1| similar to
35 in 92 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34490.1"
/db_xref="GI:13360527"
/translation="MLPILWPSARDLRLQITTYIAKENPPAARRLKIRIETSVLPIS
EHPYLYPPSPRVSGLREIVTHPNYILYRVAASSIVSVTHSRROFPFSI"
complement(30577. .30768)
/gene="ECs1068"
complement(30577. .30768)
/gene="ECs1068"
[notes="unknown"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34491.1"
/db_xref="GI:13360528"
/translation="MGTALSPVSEFETTEQENSYNENLRTKVTSSLAOTRPAIPHDE
VNAENENLIAQTAVTNKSE"
complement(30796. .31173)
/gene="ECs1069"
complement(30796. .31173)
/gene="ECs1069"
[notes="probable regulatory protein, similar to probable
regulatory protein [Salmonella typhimurium]
gii17467281|pir||IT03008 percent identity 30 in 108 aa, also
similar to bica [Escherichia coli]]
```

```
gii118631|sp|P06966|DICA_ECOLI percent identity 27 in 108
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative regulatory protein"
/protein_id="BAB34492.1"
/db_xref="GI:13360529"
/translation="MYRIKLLQETGWSQAEARRIGVTOQTVOQWVGKATPKASSL
DKLVEYSGHPLHWFLLPPECEQIFTPDTMKIGPRQRELIQAFSAFPEEQKMLQEI
KDKKKSMEETIARWLAQAQSRRA"
31306. .31578
/gene="ECs1070"
31306. .31578
/gene="ECs1070"
[notes="unknown"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34493.1"
/db_xref="GI:13360530"
/translation="MLDSTFEKIROKYTOAEIGRYMGVAOOTVWOMFSEGVGPKQVIP
LCOLMKWEVTPHPIREDIIPNPTDGLPVGCKVWTSNAPELIHENQA"
31562. .31987
/gene="ECs1071"
31562. .31987
/gene="ECs1071"
[notes="unknown, similar to ydaT [Escherichia coli]]
gii13025103|sp|P76064|YDAT_ECOLI percent identity 31 in 141
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34494.1"
/db_xref="GI:13360531"
/translation="MKIKHEHIRMANNAWAHPDGEKVPAAKITKAYFELGMTFPELYD
DSHPEAMARTQKIFRWVEKDTDAVKKIQALLPAIEKAMPPLVARMRSHSAYFRE
LVETKERLYKIDIDFVASAIVLFDQNRGGPAGNTLAVH"
complement(32194. .32649)
/gene="ECs1072"
complement(32194. .32649)
/gene="ECs1072"
[notes="unknown"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34495.1"
/db_xref="GI:13360532"
/translation="MYFLKSLYQAHVLNVAAATNRWCNSPEMLPDYRAWLRAETYLRLD
LIISLOKETASTHNLOGDAVRLIVSRHSALSIIIEVRHLISFSELIFLQPALESANI
PPEVIQYPPHVDQLQDVPYNQRAGLTPTPCSEAWDHSLLKLYQDLYNPQ"
32728. .33819
/gene="ECs1073"
32728. .33819
/gene="ECs1073"
[notes="unknown"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34496.1"
/db_xref="GI:13360533"
/translation="MRDYAKVSFRFWLGETGRELKAGAAQVVAFLYMTSPHANMLG
LVLVPLYLAHETGLLEGASGLKRAVAGFCSYDHDAMVWVHMAAWQVGETLKP
GDNRCAGVRNEYASLPENAFLSVFYDRYKTFDHLVRRNRNSRVGFGCAFGLRSQ
EQEQEQEKQEQDKNTMVHGKKNTNOAGDVQTVNPGQAGTTPPADSGAVQOVMTAG
SESHLOLQPEADSAIOREADRVVPSTQSGSVGRVDYPDVFEQWREYPLRAGANPKK
SAFSANKARLREGVPPETMLDGVRRYARYLAATGKAGTEFVQRATTTFFGPDNRFPNW
LFPVSTNQRCVNHISEPDTETPPGFRG"
33826. .34572
```

gene

```

CDS
/gene="Ecs1074"
33826..34572
/gene="Ecs1074"
/note="probable DNA replication protein, similar to DnaC
homolog [Escherichia coli] gi|7429001|pir|IC64886 percent
identity 79 in 248 aa, also similar to DnaC [Escherichia
coli] gi|118715|sp|P07905|DNAC_ECOLI percent identity 48
in 242 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative DNA replication protein"
/protein_id="BAB34497.1"
/db_xref="GI:13360534"
/translation="MKNIAAGVLERIRLAPQASVPPYTVVEWRWQIAEGRKRSE
EIRNQHLRVEKILNRSGIOPLHSCSFANYQVNDGQYALSOAKSIADELMTGCT
NFVFSCTGKGNLAAAGNRLMAGRSVYIIVSDMSVLHSDYDNGSKGEKFLQE
LCSDLLVLDEIGVQRETKNEOVVLHQIIDRTASLCSVGLMTNLNHAAMSTILGERI
MDRTWMNGGRWTFNWDSPRPVSNMRVVK"
34594..35364
/gene="Ecs1075"
34594..35364
/gene="Ecs1075"
/note="unknown, similar to gi|3025105|sp|P76066|YDAW_ECOLI
percent identity 54 in 155 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34498.1"
/db_xref="GI:13360535"
/translation="METVSDALKKASSHVVAARLGISREAVNELNELKINGVVD
KIGHTWFLAGESRYTERBPYKSEAQDMLTGEVQKVYADMMEITIGDQAKTCEEL
AKFGYSTKRVASTLAVTATGLRANQNGKFRYCPGNDLPAEPKAAALVTESDGA
FVQPAAGALPVREATAQEEIKTETVADIVQPLPSTFETQADELIFPSLRANLALRA
KSDVQKWERVCAALRELNKHRRDIVRQITDSSRRVSEK"
35380..35793
/gene="Ecs1076"
35380..35793
/gene="Ecs1076"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34499.1"
/db_xref="GI:13360536"
/translation="NAKVFTQEEERIKGVVVELVRRSGRETIRQLEAKTGATRYLMS
VLARELVASGVNNGYGLFPSEQARKDQWARKLSRAKLPVVPVDPDLIWSLPDG
EIRRYDROLNIICSECRNSEVMQRLVIFTGVNME"
complement(36145..36918)
/gene="Ecs1077"
complement(36145..36918)
/gene="Ecs1077"
/note="unknown, similar to hypothetical proteins e.g. IroE
[Salmonella enterica] gi|2738251|gb|AAC46182.1| percent
identity 29 in 249 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34500.1"
/db_xref="GI:13360537"
/translation="MTKHMIDHCOTRIFNSINHDTKYKICIFTPKNIKNNATLYLD
GNSANNYSDLIPVDALPNPVLVTLGYESWNLSIHRRAYDYTPGEGNAIVDNSKP
AWIYFTGGSGQSERELLTOIMPWSTIAPNSSRIGIMGHSIGALFVLDCLNKNSCFN
YVIYSAPSLWNERIKIKDDISESKHTKICILLNGNLSLDYSASLYPEAKAESV
LBNILTEYSNPSIVQFPELNHQETFSALMWSIIHFSI"
36960..37121
/gene="Ecs1078"
36960..37121
/gene="Ecs1078"
/note="unknown"

gene
CDS
/gene="Ecs1080"
37466..37678
37466..37678
/gene="Ecs1080"
/note="probable prophage maintenance protein, similar to
Hok/Gef family e.g. Mokw [Bacteriophage 933W]
gi|4585453|gb|AAD25481.1|AF125520_76 percent identity 92
in 65 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="prophage maintenance protein"
/protein_id="BAB34503.1"
/db_xref="GI:13360540"
/translation="MNGSKRLASYVPKGEKQAKMQLIALIVICLTIVITVLT
RDLCEVRVRTGTQTEVAVFTAYEPEE"
37846..38124
/gene="Ecs1081"
37846..38124
/gene="Ecs1081"
/note="unknown, similar to QD1 [Bacteriophage N15]
gi|2564084|gb|AAB81659.1| percent identity 31 in 64 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34504.1"
/db_xref="GI:13360541"
/translation="MAHDTKLYNSDDSAVFASRRRCFHAFKSDWYOHPPCTEEQAEW
LIQCYRRRCGVKKALSLEYRHWIISVRLPYSERPPRSRTFOQRIWR"
38198..39175
/gene="Ecs1082"
38198..39175
/gene="Ecs1082"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34505.1"
/db_xref="GI:13360542"
/translation="MSAFHNGRILVPEPKSMRALPSGVVPAVHOPLAEDKSLLPFFS
DERVTRAGAGALSDWLLRHVKSCQPHGDYHSETVIRYGTGAMVLCWCHDCNCLR
DOTSLELOAQNLAAMIDVIRHAMNGIOBRELSLAELSNWAVCNQVVDALPEAVS
RSLGLPAEKIRSVTRESDIIPGEQTATILKORTKNIALPHTHQOQNPQPEKTVVS
IAVDSPESKFMKRPKRNRVNEKYTRWYKTPQCACCGPADDPHHLIGHGGGGMGTK
SDIFTPLCREHNEHLHADPLAFBEKKGQVDLIFRFLDHAFATGLC"
39188..39559
/gene="Ecs1083"
39188..39559
/gene="Ecs1083"
/note="probable crossover junction
```

```

gene
CDS
/gene="Ecs1079"
complement(37284..37421)
complement(37284..37421)
/gene="Ecs1079"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34502.1"
/db_xref="GI:13360539"
/translation="MSLIKNLLCVFSYFIYHVDRLIYVISKTSLLFYLSTHLILI
IVYIACLGN"
complement(37284..37421)
/gene="Ecs1079"
complement(37284..37421)
/gene="Ecs1079"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34502.1"
/db_xref="GI:13360539"
/translation="MLTGAFLYLPVFMPEADSLKHPQLFYLTSHYLANMKSVRKTDV
P"
37466..37678
/gene="Ecs1080"
37466..37678
/gene="Ecs1080"
/note="probable prophage maintenance protein, similar to
Hok/Gef family e.g. Mokw [Bacteriophage 933W]
gi|4585453|gb|AAD25481.1|AF125520_76 percent identity 92
in 65 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="prophage maintenance protein"
/protein_id="BAB34503.1"
/db_xref="GI:13360540"
/translation="MNGSKRLASYVPKGEKQAKMQLIALIVICLTIVITVLT
RDLCEVRVRTGTQTEVAVFTAYEPEE"
37846..38124
/gene="Ecs1081"
37846..38124
/gene="Ecs1081"
/note="unknown, similar to QD1 [Bacteriophage N15]
gi|2564084|gb|AAB81659.1| percent identity 31 in 64 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34504.1"
/db_xref="GI:13360541"
/translation="MAHDTKLYNSDDSAVFASRRRCFHAFKSDWYOHPPCTEEQAEW
LIQCYRRRCGVKKALSLEYRHWIISVRLPYSERPPRSRTFOQRIWR"
38198..39175
/gene="Ecs1082"
38198..39175
/gene="Ecs1082"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34505.1"
/db_xref="GI:13360542"
/translation="MSAFHNGRILVPEPKSMRALPSGVVPAVHOPLAEDKSLLPFFS
DERVTRAGAGALSDWLLRHVKSCQPHGDYHSETVIRYGTGAMVLCWCHDCNCLR
DOTSLELOAQNLAAMIDVIRHAMNGIOBRELSLAELSNWAVCNQVVDALPEAVS
RSLGLPAEKIRSVTRESDIIPGEQTATILKORTKNIALPHTHQOQNPQPEKTVVS
IAVDSPESKFMKRPKRNRVNEKYTRWYKTPQCACCGPADDPHHLIGHGGGGMGTK
SDIFTPLCREHNEHLHADPLAFBEKKGQVDLIFRFLDHAFATGLC"
39188..39559
/gene="Ecs1083"
39188..39559
/gene="Ecs1083"
/note="probable crossover junction
```

```
endodeoxyribonuclease,similar to Gp67 [Bacteriophage HK97]
gi|6901639|gb|AA31142.1| percent identity 60 in 113
aa,also similar to crossover junction
endodeoxyribonuclease Rus [Escherichia coli cryptic
prophage DLP12] gi|2"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative crossover junction
endodeoxyribonuclease"
/protein_id="BAB34506.1"
/db_xref="GI:13360543"
/translation="MRIEVLPYPPVTNYYWRRRSTGYFVSKAGERYRAVALIVRQQ
RLKLSLGRLAIKITAEPPDKRRDLNLKAPLDALTHAEVLIDDEQDEINVRGQ
PVPGGRLGVIYEIRGNDGA"
39549..39920
/gene="ECs1084"
CDS
39549..39920
/gene="ECs1084"
/note="probable antitermination protein,similar to
bacteriophage antitermination proteins e.g. ybcq
[Escherichia coli cryptic prophage DLP12]
gi|4585416|gb|RAD25444.1|AF125520_39 percent identity 77
in 124 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative antitermination protein"
/protein_id="BAB34507.1"
/db_xref="GI:13360544"
/translation="KARDIQMYLERGAWAANNHEDVTPWPSIAAGFKGLIPTKVKSRP
KCSDDAMICGCMARKNNQYLHDLVDYVVGMTFEMALKRHCSDGLIGKRLYK
AEGITEGMLMALNVRDMDMR"
40072..40890
/gene="ECs1085"
CDS
40072..40890
/gene="ECs1085"
/note="unknown,similar to Ypbd [Bacillus subtilis]
gi|11730886|sp|P50730|YPBD_BACSU percent identity 30 in 128
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical membrane protein"
/protein_id="BAB34508.1"
/db_xref="GI:13360545"
/translation="MKVILVLSLTSLWHKVVAFSLTVSVVLAVLNDIIDSVLFFV
ATIVFIILKFNKNYNAKSIYEVGIVLSAIALFFHLPGFHNPVLNSVTVGQST
PYTMFNFKALVPFLVLCTLSLKFKEVSEVSLWKWGLSPLILFLAVFFGGL
KPEIHFPWLPFILLNFPVSLABESLPRGYIQSLSEVTSPLVALIIVAAALLFGFYH
YSGGALLVLVATLSGVVYGLSNWWSRLVWATLFFHGLNLCHLLFTTYPLKHN"
41177..41416
/gene="ECs1086"
CDS
41177..41416
/gene="ECs1086"
/note="unknown,similar to hypothetical protein
[Bacteriophage P27] gi|8346569|emb|CAB93762.1| percent
identity 97 in 49 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34509.1"
/db_xref="GI:13360546"
/translation="MLKQDMFETARVVEFDELSTVTEPATVGEIAQNTYLSRERCQLIL
TQLVMKWOTIISVVTDAFSPGFFTCGKWAAGGC"
41637..42224
/gene="ECs1087"
CDS
41637..42224
/gene="ECs1087"
/note="probable transcriptional regulator,similar to
transcriptional regulators e.g. Yhiw [Escherichia coli]
gi|586679|sp|P37638|YHIW_ECOLI percent identity 37 in 187
```

```
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator"
/protein_id="BAB34510.1"
/db_xref="GI:13360547"
/translation="MHYGVKIFDINHSTVSQYLEIQHKLTRHTLTDVPLYLSLEPNN
ISASWTLGAIALRLHMSSELKIKLKEGHMFSSRLLEERMRVAVNMLCSSRHGYGQAV
AEKCGSSWSYFISVFHRYGFPDPYVSRQGLDY"
42253..42328
/gene="ileZ1"
/note="anticodon: CAU, Cove score 80.94"
/product="tRNA-Ile"
42253..42328
/gene="ileZ1"
42336..42412
/gene="argN1"
/note="anticodon: UCG, Cove score 45.06"
/product="tRNA-Arg"
42336..42412
/gene="argN1"
42426..42502
/gene="argO1"
/note="anticodon: UCU, Cove score 86.06"
/product="tRNA-Arg"
42426..42502
/gene="argO1"
42992..44842
/gene="ECs1088"
42992..44842
/gene="ECs1088"
/note="unknown,similar to hypothetical proteins e.g.
[Bacteriophage 933W] gi|4585419|gb|AAD25447.1|AF125520_42
percent identity 53 in 613 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34511.1"
/db_xref="GI:13360548"
/translation="MAFKHYDVVRAASPSDLAEKLTHTKLEGMQPYGGPVAITPTYTLM
QAVAIEDGDPQVGPSSKPFDMFYVVLQAGSNMAYEGELPLPDSYDAPDPRIKQLARRS
TVTPGESCTYNDIIPADHCLHDVODMSLTNHPKADLSKGQYCGVCGOGLHIKLLPY
IPNNAGILLVPCRCGSAFTQCAEGTFSADTCASODSARWGVKPLQDLIARTKAAL
QKNPNVLLAVCMOGEFDMSAATHAQPALETAMLTFQRADLSVFNQAOCGSGSADV
PMICGDTTYWKNYATQYDVTYGVYKKNRESEGVYVPFMTDGNVNTATNAPAEQDP
IPASGYGAASRTNGNSNRPTHFSSWARRSIIPDLRLATAILNAAGTSAFISGKA
PIKTPSPGNTSPGSADTSVRTISILLPAAAGAAAGWSIKDGGIQLSDGVFKITRQS
NKTWSLTHPVDDAITLLTOGRLNCKFRLSGALTNNQFGLGIYLYTDAPVPGVAMTG
TCNPFLMSYFOTTDGRVNLHHRKAGNTKLGECGDYGNDMOTLELVFTAGSATVTPK
LNGVAGPAFQVILGSLTIGLNLTLTDVTKNAAVGEIESLVLEINAPAA"
44964..46036
/note="IS629-01, with an internal deletion"
45018..45344
/gene="ECs1089"
45018..45344
/gene="ECs1089"
/note="transposase,identical to hypothetical protein
[Escherichia coli plasmid pO157 insertion sequence IS629]
gi|7444868|pir|IT00241 percent identity 100 in 116 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase Orfa protein of insertion
sequence IS629"
/protein_id="BAB34512.1"
/db_xref="GI:13360549"
/translation="MTKNTRFSPVEVRQRAIRVMVLESQDEYDSQWRAICSIAPKIGCTP
ETLRVWVRQHERDITGGDGLTSAERKLELRENRRELRNSNDILRQASATFAKAEF
DRLWKK"
45341..45994
gene
```


gene
SPTTQIFSIICCYLKRKTNARVHHVLLLIACDLFLPHGTTFFQLEIFILQPVRIYVSS
TTLPAPPYDGPL"
complement(49003. .49209)
/gene="ECs1100"
CDS
complement(49003. .49209)
/gene="ECs1100"
/note="holin protein (host cell lysis), similar to holin
proteins e.g. [Bacteriophage VT2-Sa]
gi|5881636|dbj|BAA84327.1| percent identity 91 in 69 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative holin protein"
/protein_id="BAB34523.1"
/db_xref="GI:13360560"
/translation="MEKITGVSYTTSVGTGYWLLQLLDKVPSONWAGVILGSLLF
GLTLTLNLYEKIREDRRKVARGE"
49960. .50235
/gene="ECs1101"
CDS
49960. .50235
/gene="ECs1101"
/note="unknown,GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34524.1"
/db_xref="GI:13360561"
/translation="MATQTEVARHLSLTDRQLRRLQKLPGAPISNKRQGLDLDARDF
YISYLRKKNVDVPGDSEDDYEKLLIARWELTAEQAVTQOLKKRTA"
50223. .52660
/note="ISEC8-02, truncated (5')"
50311. .50691
/gene="ECs1102"
CDS
50311. .50691
/gene="ECs1102"
/note="unknown, similar to hypothetical proteins e.g. L0013
[Escherichia coli O157:H7 strain EDL933]
gi|3414881|gb|AAC31492.1| percent identity 99 in 133 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34525.1"
/db_xref="GI:13360562"
/translation="MOKNVTGRRKGCNPYPPEFKQOLVAASCEPGISISKLALENGI
NANLFFKRWQWREGKLLLPSESQPLLPTVLDAAEQPESLAEDPETLSICEVTR
HGTLRFNGVSEKLLTLLIQELAR"
50688. .51035
/gene="ECs1103"
CDS
50688. .51035
/gene="ECs1103"
/note="unknown, identical to hypothetical proteins e.g.
L0014 [Escherichia coli O157:H7 strain EDL933]
gi|3414882|gb|AAC31493.1| percent identity 100 in 115 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34526.1"
/db_xref="GI:13360563"
/translation="MIPLSGRTKIWLVAGITDMRNGFNGLAAKVQTALKDPPMSGHVF
IFRGRSGQVKLLWSTGDGLCLTLKRLERGRFAMPSARDGKVFLTQALAMLLEGIDW
RQPKRLTSLTML"
51085. .52623
/gene="ECs1104"
CDS
51085. .52623
/gene="ECs1104"
/note="unknown, similar to hypothetical proteins e.g. L0015
[Escherichia coli O157:H7 strain EDL933]
gi|3414883|gb|AAC31494.1| percent identity 100 in 512 aa"
/codon_start=1
/transl_table=11

gene
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34527.1"
/db_xref="GI:13360564"
/translation="MNDISSDDIFLLKQRLAEQALHALQEKLSREREDHLOAOL
DKLRMNFGRSEKSRRIAQMEADLRLOKESDITLGRVYDPAVQPLQTRTRKPF
PESLPDRDEKRLLPAAQCCPNCGGSLSYLGEDTAELMRSAFRTVIREKHACTOC
DAIVQAPASRPRIERGIAQGLLARVTSKYAEHTPLYRQSEIYVYGRGVELRRLSLG
WDACCRLLSPLEALHGYVTQDKLHADDPVQVLLPGNKKTKTGLWAVVVRDRNA
GSALAPAVFAFSPDRKGIHPQTHLACFSVLOADAVAGNELYVNGSLITEACMAHA
RRKIDHVVIRIPALTEALEQIGLYAIADIRGMPAEORLAERQKRTPLPLKLSLES
WLREKMKTLRSHSELAKAFAYALNQWPAITYANDGWIEIDNNIAENALRAVLSGRKN
FLFFGSDHGGERGALLYSLTGCKLNDVDPESYLRHLVIGVIADMPVNRVSELLPWRTA
LPAE"
52673. .52915
/gene="ECs1105"
CDS
52673. .52915
/gene="ECs1105"
/note="probable terminase small subunit, similar to
C-terminal parts of terminase small subunits e.g.
[Bacteriophage N15] gi|2507082|sp|P31061|NOHA_ECOLI
percent identity 46 in 75 aa,GTG start, probably disrupted
by IS insertion"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative terminase small subunit"
/protein_id="BAB34528.1"
/db_xref="GI:13360565"
/translation="MSKGKLIIDTGCFTALSKLAMALSSLTSLDPILSMQRQFPDLTPTR
HLDLKTLIAKAGANQCARAGDKLPDLLDEYIRATFE"
52887. .54815
/gene="ECs1106"
CDS
52887. .54815
/gene="ECs1106"
/note="terminase large subunit, similar to terminase large
subunits e.g. [Bacteriophage 21]
gi|2851579|sp|P36693|TERL_BPP21 percent identity 91 in 637
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative terminase large subunit"
/protein_id="BAB34529.1"
/db_xref="GI:13360566"
/translation="MNISEQOLNNMSAVTTALQPLTRALPVTPEWADQNYILPKRES
SYGEGEWKTLPQJAIAMNSMGNDQIRTVNLKKSARVGYTKMLGGGYYFIEHKSRNSL
LFQPTDSAAEDFMKSHVEATIRNVPCLKIDSPWLGRRKRDNTLTLLKRFSSGVGFWCLIG
GAAAKNYREKSDVVCYDELSSEFDPVEKESPTLLGDKRREGSVWPKSINGSTPKIK
GTCQIEKAANESAHFMRFYVPCPHCGEAQYLKFGDESTPFLGWKEDSPESVFLCEH
HGVVIHQSELDSQNGRCICENTGMRTDGLTFFSARGEIIPPPRSIMFHIWTAYSPT
TWQIVYDMLDALDKDPNGLKTFVNTLTIGTEWEEAVGKLDHVLMDKRVAVTAAPAR
VVILTAGIDSQRNRFENYVNGWARGEAFLVDKIIIMGRDEETLLRVDAAINKKYR
HADGETMISRVCMDTGIDGEIYIQRKKGHGFVRLPVKGSYVGRPVITMPKTRNQ
RGVYLCVGDYDTAKEILLYARMKADSPADATSYAIRFPDDPEIFSOATEAQLVAEEL
VWEKKGKMLLLNDNKKRNEALDCLVYAYAAALRVSVQRMQLDLAVLAKSREETTRP
TKELAALKSGGVNGYSR"
55002. .56594
/gene="ECs1107"
CDS
55002. .56594
/gene="ECs1107"
/note="portal protein, similar to portal proteins e.g. GP4
[Bacteriophage P21] gi|549295|sp|P36272|VG04_BPP21 percent
identity 98 in 530 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative portal
protein"
/protein_id="BAB34530.1"
/db_xref="GI:13360567"
/translation="MKRTPVLIDVNGVPLRESLSYNGGAGFGGQMAEWLPPAQASADA
ALLPALRLGNARADDLVRNNGIAANVALHKDHIVGHMFLISYRPNRWLGMRETRAAK
SFVDEVAANWSEYAEGMSGEIDVEGKRTFTFEFIREGVGVHAFNGFVQVPWDTETQO

LFRTREKAVSPKRVDTPGHGMNRFRLAGVEVDYGRAVAVHICEDDPGRSSGSRWR
IPRELTPGRPMLHIFPEVEDGOTRGANOQFYVMEBLKMLDSLOQLOSLAIVKMYA
ATIESDLTEKAFXYIAGAPOQKNDPLINILERFSWIDTNNVTGLGVKIPHLFPGD
DLKIQTODSNGSFALEALLRYIAAGLVSYEQLSRDYSKYSSARASANESWRY
FMGRKFTAAALATQMFSEYLEEALLRGIIRPRAREDFYQARSAMSRAEWIGARMA
IDGLKEVOESVMRIEAGLSTYEKEPALMGEDYQDIFRQOVESAERQKAGLSRPWIA
QAYOOIAESRRPBEETTPRET"
56584. 58089
/gene="ECs1108"
/note="ECs1108"
/note="head-tail preconnector protein, similar to head-tail
preconnector proteins e.g. Gp5 [Bacteriophage P21]
gi|549296|sp|P36273|VG05_BPP21 percent identity 97 in 501
aa,GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative head-tail preconnector protein"
/protein_id="BAB34531.1"
/db_xref="GI:13360568"
/translation="MRRNLSHIIAAAFNEPILLLPAYARVFFCALGREGMAASLSVPQ
QQVOLDAPGMLAETDEYDMAGKRPARVYRVVNGIAVLPTGTTLVRLGGMPPFSGMTG
YDGIYACLOQAMDSOVRGVLLDIDSPGQAAGAFDCADMIYRLRQOKPVWALCNDTA
CSAAMLLASGSRRLVTQTSRISGIGYMSHVYAGHLAQAGVDITILIYSAHKVDGN
QFEALPBYRNMQRIDAARRMAEKVAMFTGLSDAVTGTAAVFEFGSGDAGLA
DELVNASDAISMTALNSNVNRRGTMPTLATEAAQENQRMVILTCQEAKEGQLA
TMLAGQCMQSEYQARILAAAPQOVASTQSEADRIWACEANRQOLAATLAAMPE
MTVEKARPIIASQADAGPSLRQIINALDEAKGAQAEOLACPGMTVESARAVLA
AGCKAKBPVSASTTALFERIMANHSPPAAVGVGVPQTSDAGDADYKMLAMP"
58126. 58473
/gene="ECs1109"
/note="ECs1109"
/note="head decoration protein, similar to head decoration
proteins e.g. Gpshp [Bacteriophage P21]
gi|549437|sp|P36275|VSH_PBP21 percent identity 95 in 115
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative head decoration protein"
/protein_id="BAB34532.1"
/db_xref="GI:13360569"
/translation="MVTKNITEQRAEVRIFAGNDPAHTATGSSGISSATPALPLMLD
EAGKLVVDPGQKAGSAGVILVPLEGTETVLTYYKSGTFATEAIRWPESVDEHKKAN
AFAGSALSAAALP"
58531. 58797
/gene="ECs1110"
/note="ECs1110"
/note="possible major head protein, similar to N-terminal
parts of major head proteins e.g. Gp7 [Bacteriophage P21]
gi|547612|sp|P36270|HEAD_BPP21 percent identity 95 in 88
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative major head protein"
/protein_id="BAB34533.1"
/db_xref="GI:13360570"
/translation="MGLFTTRQLLGYTEQKVKFRALFLELFFRTVNFHTEEVMLDKI
TGKTPNAAVSPVVEGKVLKRRGGETRVLRPGYVKPKRHEPPWSR"
58914. 59519
/gene="ECs1111"
/note="ECs1111"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34534.1"

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/db_xref="GI:13360571"
/translation="MTAESYDNNYLDDDEDADWTATGQKKSAGDTSFTLAWKPGEGQ
KGLIGWESGDVRAIKRFPDNGTVDVFRGWSSICKAVTAKEVITRTVKVNVGKPSV
AEERSKITPVAIKVTPTSGTVAKGKTTTLTVSEPESATDKTRAVSADSPKATISV
KDMITITVNGVATGKVQIPVVSNGNQFAVAEVYTEAGAAG"
59461. 59964
/gene="ECs1112"
/note="ECs1112"
/note="probable tail component, similar to minor tail
proteins e.g. GpG [Bacteriophage lambda]
gi|138842|sp|P03734|VMTG_LAMB percent identity 68 in 143
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative minor tail protein"
/protein_id="BAB34535.1"
/db_xref="GI:13360572"
/translation="MVSSPQWLKSLIKRALQGRKRSYMFLEKTEQFYNGSVTLSEL
SAQRIEHLALLKRRBAQESSGNLQVSEDLVRTGAFLVAMSLWHNHPOKTPQSPMN
EAVMKIEQEVLTTPADAIARAEDVVLCLSGMSGAVRPDITITEVAKNNTLTDDDFS
GKSSTAS"
60015. 60404
/gene="ECs1113"
/note="ECs1113"
/note="probable minor tail component, similar to minor tail
proteins GpG-T [Bacteriophage lambda]
gi|7429179|pir|JTLBPTL percent identity 72 in 124
aa,probably produced by translational frameshift"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative minor tail protein"
/protein_id="BAB34536.1"
/db_xref="GI:13360573"
/translation="MLAGMTSTEYADWHRYRTHYFQDTQDLMHFSGITAVLSLFFC
DPDMHPSDFSLVPRHEEQVERPEDDKMLMQRAAGLAGGVRFGGDGRDILSSADVA
DVWVDDAALNMASAGIPGVRYVPAGH"
60385. 62964
/gene="ECs1114"
/note="ECs1114"
/note="tail length determinant, similar to tail length
tape measure proteins e.g. GpH [Bacteriophage lambda]
gi|138843|sp|P03736|VMT_H_LAMB percent identity 77 in 859
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail length tape measure protein"
/protein_id="BAB34537.1"
/db_xref="GI:13360574"
/translation="MSQVGDVLVIDSLDAVRFEQMSRVRRHFSGLDTPVRKTASAV
EGLSRQALAAQAGISVQYKAAMRTLPAQFTDIATQAGQNPWLILQLQGGQVVD
SFGCMIPMERGLAGAITLPMVGTSLAVATGALVYWGQDSTLSAFNKTLLVSGNOS
GLTDAPMLTSLRAGQAAGLTNQARSLAALVNAVGVGQFDAINOSVARFASGV
EVDKVAEAFGLTTPDPSGLIAMVROFRNVTAEQIAYVQLOSGDDEAGALQANDIA
TKGFSQTRLEKKNMGTLETWADTKGAFKSMDDAILDIGRPSSADMLASAKAFDE
ADKKWYQSRORRGKTASFRANLOGANDRENARGLAAATLQSDMEKAGELAAED
RABRDSQLKYTGEAKAYERLLTPLEKYTAQPELNKALDKGILRADYNTLMAAK
KDYESTLKPKSSGVYAGQEDQHAHALLETLELRTLEKHSANERISQQRDL
WAKENQYAVLKEAATKSEQEKFLIAHKDETLEYKROLAELGDKVEHOKRLNELAQ
QAVRFEQOSAKOAAISAKARGTLDRQAQRESEARLRDVGNDPALAKATSKLWT
WSAEQOLRGSWMAGLKGWCEWESATDSFQVKSATQTFDGIQNMAMMLTGAEND
WGGFTSVLSMLTEIFLKQAMVIGVSGAIGGAFGGGASATGTATQAAAAHFPA
TGGFTGTGKYEPAGIVHGEFVTEATSRIGVGNLYRLMRYAEGGYGGGAGSPAQ
MRRAGEINFNQNNHVVIQNDGTNGQAGPQLMKAVYDMARKGAQDELRLQLRDCGMLSG
SGR"
62961. 63290
/gene="ECs1115"
/note="ECs1115"
/note="probable tail length tape measure protein"

/gene="ECs1115"
/note="minor tail component,similar to minor tail proteins
e.g. GpM [Bacteriophage lambda]
gi|138845|sp|P03737|VMTM_LAMBD percent identity 82 in 109
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative minor tail protein"
/protein_id="BAB34538.1"
/db_xref="GI:13360575"
/translation="MKTFRWKVPDMEVNSQPSVREVRFGDGSQRYMAAGLNADLKIT
RVLSVTREARHLEAFLAEHGGWKAFLPPTPYAWRIQIKVTCAAWSRSRVMRLRVEFSA
EFKQVNV"
63290..63988
/gene="ECs1116"
63290..63988
/gene="ECs1116"
/note="minor tail component,similar to minor tail proteins
e.g. GpL [Bacteriophage lambda]
gi|138844|sp|P03738|VMTL_LAMBD percent identity 76 in 232
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative minor tail protein"
/protein_id="BAB34539.1"
/db_xref="GI:13360576"
/translation="MODHESLNSVKSQSPRVVLREIDLTVQGERYFFCNELNE
KGAVTWQGRQYQVYPIDGSGFENKGGKSARFSLTVSNLFGVLTGMAEDQLSLVGT
VVRRTVARELDVAVNFVAGPEADPOELSDRWVVEQMSLTAMTASFVLATPTETDG
ALPFGRIMLANTCMWYRSDECGYTGAVADEFDKPTTDIRKDRCSKMRGCEMRGV
ANFGGFLSINKLSQ"
64107..64742
/gene="ECs1117"
64107..64742
/gene="ECs1117"
/note="tail assembly protein,similar to tail assembly
proteins e.g. GpK [Bacteriophage lambda]
gi|139638|sp|P03729|VTAK_LAMBD percent identity 84 in 196
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail assembly protein"
/protein_id="BAB34540.1"
/db_xref="GI:13360577"
/translation="MNISAPEAYFRIAPEDWLRRAEQGEIVALVHSHPGGLPWLSEA
DRRLQKSALUSWLVCGEIHKRCVPHLTGRLFHEGVTDCTYLFRDYHLAIDMPD
FEREDWRRNGQLYLDNMEATGYRISLPSAQPGDILLCCFGSAVHAAIYCGNGE
LLHLLPEQLSKRERYSEKWQRTHSAWRHRHVSAFTGIYNDLAAASACM"
64640..65320
/gene="ECs1118"
64640..65320
/gene="ECs1118"
/note="tail assembly protein,similar to tail assembly
proteins e.g. GpI [Bacteriophage lambda]
gi|139637|sp|P03730|VTAL_LAMBD percent identity 68 in 224
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail assembly protein"
/protein_id="BAB34541.1"
/db_xref="GI:13360578"
/translation="MATTNFAFLASPPLARICLHGDLOQRCRLSLYVNTAAEIRAL
SMQPGFRRMNEGTQIRIAGDGTAPAYVYARLHEQLGEGTVIHVPRLAGAGKGL
QIVLGAIAIVGSFTAGASMALGSAAGGFSATTMLFSLGASMLILGGVAAQMLAPKA
KTPDYRATDNGRONTYFSSLDNLIAGNPNPVPYGEMLVGSRRISQDISTRDEGGDK
VWVIGROA"
65274..65480
/gene="ECs1119"
65274..65480

/gene="ECs1119"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34542.1"
/db_xref="GI:13360579"
/translation="MKAVTGRWMLSGGRHKSEKIPQCSRTGTAGALRRLSVRNYSYVT
TKTLTQGRMCRSFOGERIYRPEE"
complement(65511..66038)
/gene="ECs1120"
complement(65511..66038)
/gene="ECs1120"
/note="copper/zinc superoxide dismutase,similar to
copper/zinc-superoxide dismutases e.g. [Salmonella
typhimurium] gi|2462699|emb|CAA73588.1| percent identity
58 in 175 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative copper/zinc-superoxide dismutase"
/protein_id="BAB34543.1"
/db_xref="GI:13360580"
/translation="MKCKIIIAAMLTAASGYAAAEQVPMNLVSADCKEVSIGKITI
QETPYGILLFTPLHLSISEGIFHFVHEKGNCAPALKDKPVAALSGAGHDPKNTGKH
LQPSWDPGHGDLPALFVTHDGKANYVPLAPRLNSLKEIKGRSLMLHAGDNNHDPHE
PLGGGARMACGIIQ"
66172..69645
/gene="ECs1121"
66172..69645
/gene="ECs1121"
/note="probable host specificity protein,similar to host
specificity proteins e.g. GpJ [Bacteriophage lambda]
gi|138412|sp|P03749|VHSJ_LAMBD percent identity 65 in 1156
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative host specificity protein"
/protein_id="BAB34544.1"
/db_xref="GI:13360581"
/translation="MKGKGGRAHTPVBAKONLKSTQMMSVIDAIGEGPIEGPVKGLQS
ILVNKPTLTDNGNVTHGVTAVRAGEQQTPEGPESSGAETALGVEVTKAKPVTR
TITSANIDRLRVTFVQSLLLETSKGRNPSVRLLIQLORNGWNWTEKQVYNGKTV
SQFLAVILDLNLPREPNIHVBRETADSTSDLOKNTLWSSYTEI IDVKOCYPTAIV
GLOVDAEQFGGQOMVNVHIRGRIIQVPSNIDPEKRTYSGIWDGSLKPAISNNPACL
WDMLTPIRYGMGRKLGAADVDKWALYAIQYCDQMVDPDGGEGEPRTMFFNAYLSQQRK
AMVLSDFCSAMRCMPVNWNGQTLTFVQDSPVWVPTNSDVVDDNGGVGFYSFSA
KDRHTAVENVYTDQNGWQSTELVEDEPAILRYGRNLLKMDAFGCTSRQGAHRAGL
VITKGLLETOYDFTLGSQGLRHTPGDI IEICNDYAGTWTGGRVLSIDAASRTLTLD
REVLTPEGTAATVNLINGSCKPVSAJTAHPADRIQVSTLPDGVEYVYGVNGLPSL
RRRLFCVSTRENTDGTFFAIVAHQVPEKEAIVDNGARFEPQSGTNSVIPPVQVHLT
VEVSAADGOYLAQAKMDTPRVKGVRFSLTSGSGESRLVTATTAIDTEHRSSGLP
PGYFTLTVAINSYGOQGEPAATTFRINAPAVPATIELTPGYQITAPVRLAVDPTV
QGFWEFSETKIADTSOVETARYTGSQMSVSGPHIKPGKDFWFYVRSVNLVKGSAF
VEASGRASNDAGEYLGFLREKICKLHQAQWELIDNSQLADEMAEMKTSITETRNEI
TQVYSLDDQSAIIQIQIRVQKRDNDLALYMLVKQTKNGTIPYVAGICAGIEDTD
GQVSLNLLIADRIAMINPDGNTTPLFVAQGNLFMNDFLKLPFAVSISSANPT
FSLTPEGRLTARNADISGNVANSQTLNNVTINENCRVLGKLSANOQEGDLVKTVGKA
FPRDSRAPERPSPGTTIRVYDDQPDQRIIVPAVAFSGAKHEREHTDIYSSCRLIYR
KNGAEIYNRTALDNTLIYSGVIDMPAGHGMTLEFSVSAMLVNNVNNYPTASTISDLLVVV
MKKATAGISIS"
69713..70312
/gene="ECs1122"
69713..70312
/gene="ECs1122"
/note="probable outer membrane protein,similar to Lom
outer membrane proteins e.g. [prophage P-Eiba]
gi|7532789|gb|AAFP63231.1|AFI51091_2 percent identity 68 in
199 aa"
/codon_start=1
/transl_table=11

```

/evidence-not_experimental
/product="putative outer membrane protein"
/protein_id="BAB34545.1"
/db_xref="GI:13360582"
/translacion="MRKLYAAIILSAALCLAVSGAPAWASEHQSHTLSAGYLHVSTNVPG
SDELNGINVARYEFETLGLMVTFSFVAGDKNQLTRYSDTRWHEDSVRNRWFSVMAG
PSVRNWFESAYAMAGVAYSRVSTFSGDYLRVDNKGKTHDVLITGSDDGGRHSNTSLAW
GAGVQFNPIESVALIDIAYEGSGDWRDTGDFIVGVGYKF"
70371..71690
/gene="ECs1123"
CDS
70371..71690
/gene="ECs1123"
/notes="probable tail fiber protein, similar to putative
tail fiber proteins e.g. [Bacteriophage 933w]
gi14585436|gb|AAD25464.1|AF125520.59 percent identity 38
in 370 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative tail fiber protein"
/protein_id="BAB34546.1"
/db_xref="GI:13360583"
/translacion="MNAVKISGLVDGDKGPKVENCTIOLKARNRSATVVVNTVASEN
PDAGRYSDVEYGOYSVILLVEGFPSPHAGTITVYEDSQPGILNDFLGAMSEDDVRP
EALRRELWVEEAARHAEAKKNAGRAETSRNAGISASQAEESAAADTSAGDASES
ARQAESAASKOSEEASSSASAAKQASESSOAAEELSRKTAESAAGNAARDAT
TATEKARSAESAQSAEPIAAEAVNRITPVYPPKGEPCGAGPQGPQKDGKER
GDTGPAGATGERGPAGDAGPQGPQKGERGETGLTGNAGPQGPQKGTGAAGPAG
PQCPKGTGAAGPVGATGPGQKGPQGTGTOIRFLRGPASIIETNSHGWFPGTDGALIT
GLTFLAPKDATRVQVFFQHLQVREGDPQWDVGLDEVGSDGTGTGE"
71692..71961
/gene="ECs1124"
CDS
71692..71961
/gene="ECs1124"
/notes="unknown, similar to hypothetical protein
[Bacteriophage 933w] gi14585437|gb|AAD25465.1|AF125520_60
percent identity 93% in 129 aa, also similar to the
C-terminal part of putative tail protein of 933w
gi14585436|gb|AAD25464.1|AF125520.59 percent identi"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34547.1"
/db_xref="GI:13360584"
/translacion="MNILKKIMQRLCGCGKHDDCEHGQSULTVOLRLGPADILESDENG
IIPQDQVITQVVLDAKKIQCVVRPLQILRADGRWENIGMK"
complement(72086..72655)
/gene="ECs1126"
CDS
complement(72086..72655)
/notes="probable secreted effector protein,
similar to EspF proteins e.g. [Escherichia coli strain
E2348/69] gi12865308|gb|AAC38400.1| percent identity 37 in
87 aa, also similar to L0016 - Escherichia coli
gi13414884|gb|AAC31495.1| percent identity 38 in 126 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="EspF-like protein"
/protein_id="BAB34548.1"
/db_xref="GI:13360585"
/translacion="MAEHIPPPNPAPPPPPVQNEQSRPLPDVAQRLMQLAEHGHQIPQ
ARMAEHIPPPNPAPPPPPVQNEQSRPLPDVAQRLMQLAEHGIQPARMAEHIPPA
PNMPAPTPPVQNEQSRPLPDVAQRLMQLAEHGIQPARMAEHIPPPNPAPTPPVQ
NEQSRPLPDVAQRLMQLAEHGINTSKRS"
72631..72813
/gene="ECs1125"
CDS
72631..72813
/gene="ECs1125"
/notes="unknown, GTG start"
/codon_start=1
/transl_table=11

```

```

/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34549.1"
/db_xref="GI:13360586"
/translacion="MPEYVQPTSGLYECHALQDAAPDAELHQVEADVKRFFLYLITEY
KMDVVELIILYKNFLP"
complement(73294..73476)
/gene="ECs1127"
CDS
complement(73294..73476)
/gene="ECs1127"
/notes="unknown, similar to part of avirulence protein A
[Pseudomonas syringae] gi1147261|sp|P11437|AVRA_PSESG
percent identity 46% in 56 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34550.1"
/db_xref="GI:13360587"
/translacion="MSGTSGSSDAALATRYAAEYFCKTWTAFGLNOAEGYKAISDLS
HYFRABEGSSPPQSFL"
complement(73581..73668)
/gene="serT"
/notes="anticodon: UGA, Cove score 76.89"
/product="trna-Ser"
complement(73581..73668)
/gene="serT"
74094..75212
/gene="ECs1128"
CDS
74094..75212
/gene="ECs1128"
/notes="similar to HYAA_ECOLI gi11787206 percent identity
99 in 372 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hydrogenase-1 small subunit"
/protein_id="BAB34551.1"
/db_xref="GI:13360588"
/translacion="MNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKTAW
ALENKPRIPVWIHGLECTCTESFIRSAHPLAKDVTLSLDYDDTLMAAGTQAE
EVEFDIITQYNGYILAVEGNPLPGEQMFICISSGRPFIEKLKRAAGASAIITWGTG
ASWCGVOAARPNPTQATSIDKVIDRPIIKRGPPPIPVMSALITITVTFDRLPDVD
RMGRPLMFYQRIHDKCYRAHFDAFEGVQSDDDAARKGYCLYKMGCKGPTTYNACS
STRWNGSVFPIQSGHGCGLGCAENGFWDRGSFYSRVVDIPQMGTHSTADTVGLTALGV
VAAAVGVHAVASADVQRRRHNOQPTETEHQPGNEKQA"
75209..77002
/gene="ECs1129"
CDS
75209..77002
/gene="ECs1129"
/notes="similar to HYAB_ECOLI gi11787207 percent identity
99 in 597 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hydrogenase-1 large subunit"
/protein_id="BAB34552.1"
/db_xref="GI:13360589"
/translacion="MSTOVETOCYTINNAGRRLVDPITRIEGHMRCEVINDONVIT
NAVSCGTMRGLEIILQGRDPRDAWAFVRICGVCVGHALASVYAEIDAIGIKVDPN
ANIRNIMATLMDHDLVHFYQAGMDVIDLDALADPRKTSYSELAQSLSSWPKSSP
GYFVDONLKFVEGGLGIPRNGYVGHQPYQLPEANLMGFHYLEALDFQREIVK
IHAVFGKPNHPNWIYVGGMPCINIDESGAVGAVNMRLNVQSIITRTADFINNNVI
PDALATIQNKWSEITGTLSKCVLSYGAFPIDANDFGKSLMLMPGAVINDFNVI
LPVLDVDPQVOEFVDHWARYPNDOVGRHPDPDITDPWYNGDVKSGSDTNIQOLNEQ
ERYSWIKARPRGRNAMEVGPLARTLLAYHKDAATVESVDMMASNLNPLSGIQTSLG
RLICRAHEAQAAGKQYEFNKLMTLKNGLATATKEWPTETPECKRGVGTETAP
RGALGHWAARIDGKIDLYQCVPVPTWNASPRDPKQIGAYEALMTKWAIPQPLEI
LRTLSHFDPLACSTHVLGDGSELISVQVR"
77021..77728
/gene="ECs1130"
CDS
77021..77728
/gene="ECs1130"

```

```

/note="similar to HYAC_ECOLI gi11787208 percent identity
100 in 235 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="probable Ni/Fe-hydrogenase 1 b-type cytochrome
subunit"
/protein_id="BAB34553.1"
/db_xref="GI:13360590"
/translation="MOOKSDNVVSHYVFEAPVRIVHMLTVLCMAVLMVMTGYFIGKPLP
SVSGEATYLFYMGYIRLIHFESAGMFTVLLMRIYWAFVGNRYSRSELFVPPWRKSMW
QGWVEIRWYLELAKRPSADIGHNPQAAMFCYFLMSVFMITGTALYSEHSQYALF
APRYVVEFYWTGNSMDIHSWHRLGWLIGAFVIGHVIMALREDINSDDTIVSTMV
NGYRSHFKGKISNKNERS"
77725..78312
/gene="ECs11131"
CDS
77725..78312
/gene="ECs11131"
/note="ECs11131"
/note="similar to HYAD_ECOLI gi11787209 percent identity
98 in 195 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="processing of HyaA and HyaB proteins"
/protein_id="BAB34554.1"
/db_xref="GI:13360591"
/translation="MSEQRVVMVGLNLLWADEGFCVRAERLYAHYHWPESVEIVDG
GTGLNLLGVESASHLLILDADYGLPEGTLTTRYAGERIPAYLSAKKSLHONSFE
VIALADIRGLPHALIALVGLQPAMLDDGSGISELAREQLPAAEQAAQLAANGIVP
QANESRCUNYDCLSMENIEGVKLRQYRMTLEQQ"
78309..78707
/gene="ECs11132"
CDS
78309..78707
/gene="ECs11132"
/note="ECs11132"
/note="similar to HYAE_ECOLI gi11787210 percent identity
99 in 132 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hydrogenase 1 formation factor HyaE"
/protein_id="BAB34555.1"
/db_xref="GI:13360592"
/translation="MSNDTPFDALWQRMILARGWTPVSEFRLLDDWL/QAPDGVVLLSSD
PRTPVSDNPVMIGELREFPDYTQWVAIADLEQSEAICDRGVRFRFPATLVFTGGN
YRGVLNGIHPWAEILNLMRGLVPEQERAS"
78704..79561
/gene="ECs11133"
CDS
78704..79561
/gene="ECs11133"
/note="ECs11133"
/note="similar to HYAF_ECOLI gi11787211 percent identity
99 in 285 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hydrogenase-1 cofactor biosynthesis protein
hyaF"
/protein_id="BAB34556.1"
/db_xref="GI:13360593"
/translation="MSETFHLGPRTPQNDDFSMNPLPITCQVNDPESMAALEQCA
HSPQVIALINELOHQSRQPPGLGEVLAVDLLNADDRHFINTLIGEVEVSRIQQA
DSESIQIABIFCGLWVRRRRKEKLEDKLEAGCAPLALWQAATQNLPLTDSLLPP
IDGLMNGPLPLAHLAHVRNPQAQPHSINTQPISEADRLFLSRCLGPGNTQIRTIG
YGESYINATGLRHVHLRCTDITLKGPLLESYEICPIPEVVLAAPEDLVDSAQRLSEVC
QWLAEEApt"
79695..81239
/gene="ECs11134"
CDS
79695..81239
/gene="ECs11134"
/note="ECs11134"
/note="similar to APPC_ECOLI gi11787212 percent identity
99 in 514 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental

```

```

/product="probable third cytochrome oxidase subunit I"
/protein_id="BAB34557.1"
/db_xref="GI:13360594"
/translation="MWDVIDLSRWQFALTALYHFLFVPLTLGLIFLAIMETIYVVYTG
KTIYDMTRFMCKLGCINFAIGVATCLTMEFOGTNNSEFYNVGVGDFGAPLAMEALM
AFLESTFVGLFFQWQRLNKYQHLLLVTLVAFGSNLSALWILNANGWQYPTGAHFD
IDTLRMEYTSFSELVFNVPVSKVFTVWAGYVGTGAMFIMASWYLLRGREKDVLR
SFAIGSVFGLTAIIGTLQIGDSSAYEAQVPVKLAAMEGEWQTEPAPAFHVVAWPE
QQERNAFAIKIPVLLGILATHSLDKPVLGLNMAETYPRLQGRMAWLLMQEISQG
NNEPHVQLAFRELEGDLGYGSMULSRVAPDMNHVTAAYQAAMRGAIPQVAPVFWSPRI
MYCGSLLILLVMIJALVOTLRKIDOHRAVLMKAWLSLPLPWIAIEAGFMTEFEGRQP
WAIQDILPHYSAHSALTTCQLAFSLINIVGLITFLTFLIAEYVLMOKYARLGPSNQSEQ
PQQQQ"
81251..82387
/gene="ECs11135"
CDS
81251..82387
/gene="ECs11135"
/note="ECs11135"
/note="similar to APPB_ECOLI gi11787213 percent identity
99 in 378 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="probable third cytochrome oxidase subunit II"
/protein_id="BAB34558.1"
/db_xref="GI:13360595"
/translation="MPDYETLRFIMWLLIGVILVFMISDGMFGICLLPLVARDDD
ERRIVNSYGAHWEQNOVWILLAGALFAAPRVYAAAFSGFYVAMILVCLSLFFRPL
AFDYCKIADARWKMDAGLVGSLVPPVVGIAFGNLLGVPPAFTPOLRVEYILGS
FWQLTTFPFLCGLLSLGMVILQGGVWLQIKTVGVHLSQSLSTRKRAALLVCLFLLA
GWLWVGIDGVLQAODANPNLMKVLAVLPAGMNNFVESPVLWIEPLIGFCFCLP
LTVMAIYRGPRGWMGLMASLMQFGVIFTAGITLFFPEVMPSSVSPISLTLMDSTSQL
TLSIMLVILFIFLVILVLYTLWSYKMGWMTTETLRRNENELY"
82572..83876
/gene="ECs11136"
CDS
82572..83876
/gene="ECs11136"
/note="ECs11136"
/note="similar to phosphoanhydride phosphorylase
APPA_ECOLI gi11787215 percent identity 98 in 434 aa
(Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="phosphoanhydride phosphorylase"
/protein_id="BAB34559.1"
/db_xref="GI:13360596"
/translation="MKAILIPLSLLIPTOSAFQAQSEPELKLIESVVIYSRHGVR
APTKATQLMDQVTPDAWNPVVKLGLWLTFRGGELIAYLGHYQRORIVAGLLTKKGP
QPGQVAIIADVDERTKTEGAAGAIPDCAITVHTQADTSSPDPLFNPKTGVCOLD
NANVTDAILSRAGSGIADFTGHRQTAFRELYVLPQSNLCNLRKQDECSLTOAL
PSBLKVSADNVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDSHOWNTLISLHNA
QVYLLQRTPEVARSRATPLDLIMIALTPHPQOKQAYVLTPTSVLFIAGHDTNLANL
GGALELWTLPGQPDNTPPGELVFERWRRLSDNSQWIOVSLVFTQLQMRDKTPLSL
NTPPGVKLTLGACCEERNAQCMCSLAGFTQIVNEARIPACSL"
complement(83996..86176)
/gene="ECs11137"
CDS
complement(83996..86176)
/gene="ECs11137"
/note="ECs11137"
/note="similar to YCCC_ECOLI gi11787216 percent identity
99 in 726 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34560.1"
/db_xref="GI:13360597"
/translation="MTTKNMNTPPGSQNEIDLRLVGLMDHRRKFIISVTLFTLI
AVAYSLLSTPIYQADTLVQVEQKGNALISGLSDMIPNSPESAPIQLIQRMILGK
TAEILNRDMVEQKYPPIVGRWARLTKKPGELAISWMHIPOLNQDQDLTLTVGEN
GHVTLGEETVNGVGRLEKDGVALTTADIAKAPCTOVLRSORLEAINALQETF
TVYSERKESGMLELNTWGDGDPQLITRILSIANNIYQQNIARQAQSDSLSLELRQL
PEVRSLEDQAEKLNVIQRQSDVDLLEAKVLEQIVNDQNLNLTIFREAEISQLY
KKDHPYTRALLEKRLQTEQERKLNRKVPKAMPSTQOEVLRLSRDVBAGRVYQLLNR
QOELSTKSKSAIGNRIIDPAVTQPPQVPRPKKALNVVLGILGLFTISGAVILARMLR

```

RGVEAPEQLEBHGISVYATIPMSEWLDKTRRLRKKNLFSNOORHRTKNIPIELAVDNPA
DSAVEAVRALRSTLSHFAMMETENILMTGATPDGSKTFYVSSSTILAAVIAQSDQKVLFI
DADLRGYSYHNLPYVSNHGLSEYLAGKDELNKVIQHFGGDFVITTRGQVPPNPSEL
LMDRMKQLLEWANDHYDLVITDPPMLAVSDAAVGRSVGTSLVARFGLNTAKESV
LSMRQLKQSWNTKGAAILNGVYIKRSTAYSISYNYGYISYSEKE"
/gene="ECs1138"
/complement(86196..86654)
/gene="ECs1138"
/complement(86196..86654)
/note="similar to YCCZ_ECOLI gi|1787217 percent identity
99 in 152 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative phosphatase"
/protein_id="BAB34561.1"
/db_xref="GI:13360598"
/translation="MDSNMAQLKFNLSILVCTGNICRSPIGERLLRKLRLPGVKVSAG
VHGLVKHPADATAADVAANHGSVLEGHAGKRLTAEMARNYDLILAMESEHIAQVTAIA
PEVRGKTMFLGOMLEQKEIPDPYRKSDAFEHYVGMLEASQEWAKRLSR"
/complement(86630..87769)
/gene="ECs1139"
/complement(86630..87769)
/gene="ECs1139"
/note="similar to YCCZ_ECOLI gi|1787218 percent identity
100 in 379 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative function in exopolysaccharide
production"
/protein_id="BAB34562.1"
/db_xref="GI:13360599"
/translation="MKKNIFSVLTVLAVLSLRACTLVPQONLSTSKNOVIELPDNQY
DLDMQNIYVPTGLIDQLRAKIPMSOANPELQEQIANYEYRIGDVLMTVMDHPE
LTPPAGQYRSEFTSGMNVNADGAIFYPIGRKLAVAGTKLTQVRNEITARLDSYDIEPQ
VDVSVAAFRSQKAYVTGEVSKGQOPIITNPLTMDAINAAGGLTADADNRNVLTGN
GVKTKNLYALMGDLRQNKLLPGDILFIPRNDLKVFMVGMCEKQSTLKKDRSGM
TLAALGNAGSMQDVADATGFIIVRATONKONGKATANIYQLNAKDSAMILGTEFQL
EPDVIYVTPAPLARNRVVISLVLPTGSLGVHDLTETSRWIQTWPN"
/complement(87815..89911)
/gene="ECs1140"
/complement(87815..89911)
/gene="ECs1140"
/note="similar to YMCA_ECOLI gi|1787219 percent identity
99 in 698 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34563.1"
/db_xref="GI:13360600"
/translation="MKKNISYLLSCLAIVSACHAEVLTYPDPLGSSQSDFGGTGLLQ
MPNARIAPGEFVSVDNDQYRFSYTSVALFPWLEGTIRYTDVRKYQSOWDFDSDG
QSYKDFSPFKRLWEQYKLVPOVAFGRKDIAGTGLFDEGYLVASKQAGPEDFTLGMA
WYAGNAGNTNPFCRVSKYCHRAESHADGDISFSDIFRGPASIFGGIEYQTPWNP
RLKLEADGNNYQNDPAGKLPQASHFNWAGVYRAASWADLNLSYTERNTLMFGTLRTN
FNLDRPALRDTPKAPQAPESGLQYTTVAQNLTKYNAGDAPEIQLRDKTLYS
GOQYKYDRSEAVDRANRILVNNLPQVEKISYVTKREHAMVTTETDVASLRKQLAG
TAPQSEPIQQQORVEADLSAFGRYIRREDRFSYFNPTLSQSLGGPEDFYMFOLG
MSARYWFDTDLHLLDGGIFTNINYNDYDKFKSSLPADSTLPVRVTHIRDYVRNDVYLN
NLQANYFADLGNQYGVQVYLETWYAGVGSLELLYRPLDASWALGDVYNYKQRDWD
NMRFDTSTPFGFVAYNNPTLNGVLMLKSVQYLAKDKGATIDVARFDSGVAVG
VWAASVNSKDDYGGEGFGKGFYISIPFDLMTIPGNRNRNAVSWTPTLTRDGGQMLSRK
YOLYPMPTAEREVPVQ"
/complement(89911..90657)
/gene="ECs1141"
/complement(89911..90657)
/gene="ECs1141"
/note="similar to YMCB_ECOLI gi|1787220 percent identity
98 in 248 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11

/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34564.1"
/db_xref="GI:13360601"
/translation="MNKLOSIFYASVLYVWTPHAFAGQVTIYILPGEOOTLSVGVPEN
VAOLVTPQLRDLRWPFGALLTDSAVKAKALKDQYHWAQLASWEAADDVAATIKS
VROQLLNLTGRFLPDLDFVRVDENSNPPLVGNNTYTLTVORPVTITLLIGAVSGAG
QLPMQAGRSVTDYLODHPRLAGADKNNVITVITPEGETVWAPVALWNKRHRVPEPPGSQL
WLGFSAHVLPKPKYADLNDQIVSVLTQRVP"
/complement(90654..91298)
/gene="ECs1142"
/complement(90654..91298)
/gene="ECs1142"
/note="similar to YMCC_ECOLI gi|1787221 percent identity
98 in 214 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative regulator"
/protein_id="BAB34565.1"
/db_xref="GI:13360602"
/translation="MRPLILSIFALFLAGCTHSQOSMVDTFRSLFDNQDITVADQOI
OALPYSTMVLRNLNEGQRFVVLGYIEQESKWLSDQNMALVTHNGRLKTKVKNLNL
EVTNSGQDPLRNALAIKDGSRWTRDILWSENHFRSATLSTSFAGLEHLHIAGRDV
LCNVQBEVSTLPERQWNTFWYDSATGQVROSQMLGAGVIPVEMFLKPAD"
/complement(91405..91728)
/gene="ECs1143"
/complement(91405..91728)
/note="similar to YMCD_ECOLI gi|1787222 percent identity
98 in 107 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34566.1"
/db_xref="GI:13360603"
/translation="MNKGVKMHKLSAILMAFMMLTTPAFAAFAAANGTEATTGTCTT
TTTTTATTATTGGVAGAVGTATVGVATAGVATLAVVAANDSGDGGSHNTSTT
TSTTR"
/complement(92152..92364)
/gene="ECs1144"
/complement(92152..92364)
/gene="ECs1144"
/note="identical to CSPH_ECOLI gi|2367112 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="cold shock-like protein"
/protein_id="BAB34567.1"
/db_xref="GI:13360604"
/translation="MSRKMTGIKVTDFDKSGKGIIPSDGRKEVQVHISAFTPDRAEV
LIPGLAVECRVNLGRGPTAANVYLS"
/complement(92650..92862)
/gene="ECs1145"
/complement(92650..92862)
/gene="ECs1145"
/note="identical to CSPG_ECOLI gi|2367114 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="cold shock-like protein CspG"
/protein_id="BAB34568.1"
/db_xref="GI:13360605"
/translation="MSNKMTGLVKWPNADKGFITPDDGSKDVFVHFTAIOSENPRT
LNEKNQVEFSIEQGQGPAAANVYTL"
/complement(93036..93266)
/gene="ECs1146"
/complement(93036..93266)
/gene="ECs1146"
/note="similar to SFA_ECOLI gi|1787225 percent identity 96

in 76 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Sfa protein"
/protein_id="BAB34569.1"
/db_xref="GI:13360606"
/note="similar to MHRWISQNNIRLPCGAFFISVLFFFNACVIVSNLLIIIESFGM
AYNISYTRVPGTNTLLACCLLRPEVNSEY"
complement(93477..94550)
/gene="ECsl147"
/complement(93477..94550)
/gene="ECsl147"
97 in 357 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34570.1"
/db_xref="GI:13360607"
/translation="MAEKKRTRQRRPGTTGGKLPWNDWHNATTWRKATOLLILAMNI
XIAITFWYVYETAGSTIFVARPGIEGHLPIAGLMNLKYSLATQQLSVHAAAML
LLVAFIVISLLKAKFCSWLCPVGTLSLSELGALGNKLFGRQCVLPRLWDIPLRGVKYL
LLSFFLYIALMLPAQAIHFMSPYSVMYDKMLDFFRHGTTATLISVILLIASLFI
RHAWCRYLCPYALMGVYSLSPFKIRRNAESCIDCGKCAKNCPSRIPVDKLIQVTV
ECTGCMTCVSCPVASTLTFSLQKPAANKAFALSGWLMTLLVLGIMFTVIGYAMVAG
WVOSPVEELYRLIQAQPMIGH"
complement(94622..97336)
/gene="ECsl148"
complement(94622..97336)
/gene="ECsl148"
/note="similar to TORS_ECOLI gi|1787227 percent identity
97 in 904 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="sensor protein tors"
/protein_id="BAB34571.1"
/db_xref="GI:13360608"
/translation="MGFALMALLTSTLVGWNLRFIISQVEKDNTOALLIPTNMNARO
LSASAWELFAQNLTADNEKMQAGRMLTQSLKINALQALRQGGEDTTAIEQQ
EQEISLRQOQGLVGRRLQRLQSOQIVAAADEIARLAQOQANNAATAGATQ
AGYLDTEQOQAAESALDRIDIDLEYVYNMNEELRIALRQOQMYNMLGLEIQKN
APTLEQNNNAVKILQRQRIEDPGVRAQVATLTTVSOYSDLLALFOODSEISNHL
QTLAONNIAQFOAFSEVSOLVDTIELRNQHGHLAHLKASARGOYSLLLILGMYSLCAL
ILLIRWVYSVTPELAEQTLQRLDGDIDSPFETAGVRELDITGRLDAPRSV
HALNRHQLAQVAKTAELQELVIEHRQARAEAKASQAKSAFLAAMSHEIRTPLY
GILGTALLADNPALNAQRDDLRAITDSGESLITLNDILDYSAIEAGKNVSYSDPP
FEPRLLESTLQMSGRVKGRPIRLATAIADVPYALMGDPERRIRQVITNLLSNALRF
TDEQIVLRSTDEQEWLVEVDSGGCIDPAKLAEIFQPTIOVSKRGGTGLGITISS
RLAQMGGELSATSTPEVSGFCFLRLPLRVATAPVKTQVNOAVRLDGLRLILLIEDNPL
TORITVEMLTSGAQVVAIGNAQALETLONSEPPFAAALVDFDLPVNGITLQRLAR
QYPSVLIGESAHVIDETLKRQTSLSLPPIPKPVPREVLGQLLAHLYLQLOANDLPL
DVSLNDEDAQLMGTEKTHEWLAFTQHALPLDEIDTARATQDSEKIKRAAHLKSSC
SSLGMRSAQQLCAQLEQOQLPSALPHBEITRSVAALEAWLNKKDLNAI"
97449..98477
/gene="ECsl149"
97449..98477
/gene="ECsl149"
/note="similar to TORC_ECOLI gi|1787228 percent identity
99 in 342 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="periplasmic protein TorT"
/protein_id="BAB34572.1"
/db_xref="GI:13360609"
/translation="MRVLLFLLSLFMLPAFASADNLLRWHDAOHTTVOASMPLKAKRV
WKLCALYPSLKDSYWLISLVGMQEAARYGVDLKVLKLEAGCY SOLATAQOQIDOCKWNG
AEAILGSSITSFPDLQKQVANIPLVIELVNAIDAPQVKS RVGVPWFQNGQPGRYLVQ
WAHGKPLUNLLMPGPDNAGGSKWEGFRAAIAGSPVRIVDIALGDNDIDIEIQRNLQGE
MLERHPIDIVVAGTAIAAEAMGEGRNLTPLTVVVSFYLSHQVYRGLKRGVIMAASD

QMWWQELAVEQAIRIQGQSVSDNVSPPIVLTPKKNADREHIRRSLSPGGRFVPVY
QHTSAAKK"
complement(98450..99142)
/gene="ECsl150"
complement(98450..99142)
/gene="ECsl150"
/note="similar to TORR_ECOLI gi|1787229 percent identity
99 in 230 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="response transcriptional regulator for torA"
/protein_id="BAB34573.1"
/db_xref="GI:13360610"
/translation="MPHHIVILVEDEPVTQARLQSYFQEGYTVSVTASGAGLEIMON
QPVLDILLDINLPDENGLMLTRALRSTVGIILVTRGSDRIDRIVGLEMGADDDYTKR
PLELRELIVRVKNLWRLIDLARQAQPYTDONCYFCAGYCLNVSRRHTLERDGEPIKLR
AEYEMLVAFVTPNGEILSRERLLRMLSARRVENPDLRTVDVLRRLRHLKLSADLLVTO
HGEGYFLAADVC"
99272..100444
/gene="ECsl151"
99272..100444
/gene="ECsl151"
/note="similar to TORC_ECOLI gi|1787230 percent identity
99 in 390 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="trimethylamine N-oxide reductase cytochrome
c-type subunit"
/protein_id="BAB34574.1"
/db_xref="GI:13360611"
/translation="MRKLWNLRRPSARWSVLALVIGIVIGIALIVLPHVGKIVTST
TEFCVCSCHSQPVYEEYKQSVHEQNASVRAECHDCHIPDIPGMVVRKLEASNDIYQ
TFTAHSIDTPEKFEAKRAELAEARWAKMKNNSATCSCHNYDAMDHAKQHPAAROM
KVAADNQSCIDCHKIAHQLPDMSSGRKOFDELRAAGDSGDTLYSIDIKPIYAAK
GDKEASGLLPASSEVKYLRKRDQWLQIEITGWTESAGRQVLTQFPKGRIFVASIRGD
VQQQVTKLEKTVAADTNTWMSKLOATAMKMGDMVNDIKPTIWAYADSLYNGTCNOCHG
APEISHFDANGWIGTLLNGMIGFTSLDKREERTLLKYLQMNASDTAGKAGHKKEK"
100444..102990
/gene="ECsl152"
100444..102990
/gene="ECsl152"
/note="similar to TORA_ECOLI gi|1787231 percent identity
99 in 847 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="trimethylamine N-oxide reductase subunit"
/protein_id="BAB34575.1"
/db_xref="GI:13360612"
/translation="MNNDLFOASRRRLAQGLGLTVAGMLGPSLLTSRRATAQAAT
EAVISKEGILTSHWGAIKRAITVRKDVRAAKPELDKYPSKMTAGLPDHVHNAARIY
PMVVDMLRRKRLSDTSQRGDNFRVRVDEALDMFYEEELVKQTKHGTGHPVHNAARIY
QOSTGMFNASGMKAKAIQALHNSVGTGDTSTGAQVILPRVYSGMSEVIEQQTSWMLY
LQNSKTTVLGMSDLLKKNQAMWCPDHDVYEYTAQLKAKVAAGEIVISIDPVVTSTG
EYLGRHVKHIAVNPQDVLQALAHLYSENLYDKNFLNACVGFQEQLPYLLGSK
DGQPKDAWAEKLTGIDAETIRGLAROMAAARTQIIAGWCVORMHCEQWAMTVVIA
AMQLQICLPGGFGFGHHYNGAGCTPGKRGVILSGFSGSTSIIPVVDHNSDYKGYSTIP
IARFDAILLEPGKVINWNSVKVLPPLKMCIFAGTNPFRRHQINRIIEGWKRLKLETVI
AIDNQWTSCTCFADIVLPATTPQFERNDDQYGNHNSGIIAMKQVVPQPFARNDPDI
FRELCRRFRNREAPTEGLDEGMWLKRIWQEGVQOQKGRGVHLPADFDDFWNNKEYVERD
HPOMFVRHOAFREPDPLEPTSGLTIEYSKTIADNYDDCCQHPMWFKEKIERSHGG
PGSQTYPLHLQSVHPDFRLHSQLESETLRQOYTVACKPEVFINPDASARGIRNGDV
VRVFNARGQVLAVNSDYRPAVARIHGAHDPDKGGEFGALCKGNPNVLTIDIG
TSLOAATSAAHTTLVEIEKCNGTVEQVTAGVPMVMAQCEYVPASQVKL"
102987..103586
/gene="ECsl153"
102987..103586
/gene="ECsl153"
/note="similar to TORB_ECOLI gi|1787232 percent identity
96 in 199 aa (Conserved in E.coli K-12)"
/codon_start=1


```
gene
CDS
/product="putative excisionase"
/protein_id="BAB34584.1"
/db_xref="GI:13360621"
/translation="MBELYNQHNHG1QPVTIPVVOINANEWVTELLMAVTGLRKGTI
LRARDASWNGRYKQIADGTPKKNSECLYHLPTINTWIKNQLPSQDV"
/complement(110608. .110919)
/gene="Ecs1162"
/complement(110608. .110919)
/gene="Ecs1162"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585380|gb|AAD25408.1|AF125520_3"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34585.1"
/db_xref="GI:13360622"
/translation="MEMIMACSTFNPLTLQKYQDPDEDLCSLCGGNHGKAAMTECKDK
IH1CLNCVDVLVDIKNEREDKRSEAVRALDSWMRDGYSAAQIYDLAISKEIPGVRI
E"
/complement(110979. .111323)
/gene="Ecs1163"
/complement(110979. .111323)
/gene="Ecs1163"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585381|gb|AAD25409.1|AF125520_4
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34586.1"
/db_xref="GI:13360623"
/translation="MVLSKLTGCRYONRRRLTVANLQAVKGEYFDAMIGKTEEE
YRLPNYWKRMFREYDRLIITKGYPKRDSRRIDVPYDGEIKTITTHPHFEDRPV
KVFAIKVNIQNE"
/complement(111256. .111879)
/gene="Ecs1164"
/complement(111256. .111879)
/gene="Ecs1164"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W]
gi|4585382|gb|AAD25410.1|AF125520_5, similar to
hypothetical protein [Bacteriophage 933W]
gi|4585455|gb|AAD25483.1|AF125520_78 percent identity 50
in 80 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34587.1"
/db_xref="GI:13360624"
/translation="MTFTFDKMIKEIKERIGSLDVRDNIERRAYEIALTALTEPFA
TIDTVGIELVKYCNFTICPDNSMEPGNVPLYIGLPRIDPASQTAKLSQEWLSEQKE
KIDVDCGVSIETLTHWMSAYEAGNSVPTDWSICSDRMEKPGQNVLIISNFDSSL
VEPLCSARYTGSTGFRRGDATIKPGNGIEQATHWMLPEPPQEVNRG"
/complement(111883. .112170)
/gene="Ecs1165"
/complement(111883. .112170)
/gene="Ecs1165"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585383|gb|AAD25411.1|AF125520_6
percent identity 100 in 95 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34588.1"
/db_xref="GI:13360625"
/translation="MATIQELIDLTPQEKANRLKAVKDFRAAGKGYFVLDLTLSA
YNGEHVASINDKGYHTASVIMPESIDAPGLTSWADDWHGITLKDGVGVKDK"
/complement(112172. .112390)
/gene="Ecs1166"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585384|gb|AAD25412.1|AF125520_7
percent identity 100 in 72 aa, GTG start"
/complement(112172. .112390)
/gene="Ecs1166"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585385|gb|AAD25413.1|AF125520_8
percent identity 100 in 95 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34589.1"
/db_xref="GI:13360626"
/translation="MPTLFRKEYPRKSRAFEFLILFIVLMTPTISPLIFVMAIGKII
ELVIELYNDVWASFTLHNKINPYKEN"
/complement(112392. .112679)
/gene="Ecs1167"
/complement(112392. .112679)
/gene="Ecs1167"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585385|gb|AAD25413.1|AF125520_8
percent identity 100 in 95 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34590.1"
/db_xref="GI:13360627"
/translation="MVFAKSPARSKTAGTTTCALKESDMAIAASYTMHLYCDCLQCTD
GKYKSPDFGEYIGTSHAGCAKEARKDGRISRDKTRAFAPGHKILRSNKEG"
/complement(112609. .113076)
/gene="Ecs1168"
/complement(112609. .113076)
/gene="Ecs1168"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881600|dbj|BAA84291.1| percent
identity 100 in 155 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34591.1"
/db_xref="GI:13360628"
/translation="MLRPTLIAMHRLTKSLWRFTSGFGNKVPSAPLAPASKEVSNA
CGMYRLVTVSDFLGCANCCKEHHPDTPKAPPYCSGTTSSKEMQESKTMKRRKAKIL
LVRNAPGVQWVRLSNRRRMGLMKYIGMMDCGFCCKPQSAQNWKHNLRTKGE"
/complement(112949. .113722)
/gene="Ecs1169"
/complement(112949. .113722)
/gene="Ecs1169"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585386|gb|AAD25414.1|AF125520_9
percent identity 100 in 257 aa, similar to hypothetical
proteins e.g. [Bacteriophage 933W]
gi|4585455|gb|AAD25483.1|AF125520_78 percent identity 95
in"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34592.1"
/db_xref="GI:13360629"
/translation="MSEINYQALREKAERKATGSIYVGHTSVNHQNLGTGVFCQKWK
GEPGVIAECHVNCILTESDQVAEANAFTAEANPATVVALLDQERNQOYIKRRDOEN
EETALTVGKLRVELEAENNLIDSECHVALEALRDQALLSEAKRNKAKLOSENAY
IRNRYKELDLICKNILVQQAALIEWQATGDAKSLAWIYNTLFGGPELPDESEKDAQ
AYFNRYAPIDEKLMALHMKWFQESAEARAAGIRIKRGE"
/complement(113719. .113940)
/gene="Ecs1170"
/complement(113719. .113940)
/gene="Ecs1170"
/notes="unknown, similar to C4-type zinc finger proteins
(Trar family) e.g. gi|4585456|gb|AAD25484.1|AF125520_79
percent identity 79 in 73 aa"
/codon_start=1
```

```
CDS
/complement(112172. .112390)
/gene="Ecs1166"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585384|gb|AAD25412.1|AF125520_7
percent identity 100 in 72 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34589.1"
/db_xref="GI:13360626"
/translation="MPTLFRKEYPRKSRAFEFLILFIVLMTPTISPLIFVMAIGKII
ELVIELYNDVWASFTLHNKINPYKEN"
/complement(112392. .112679)
/gene="Ecs1167"
/complement(112392. .112679)
/gene="Ecs1167"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585385|gb|AAD25413.1|AF125520_8
percent identity 100 in 95 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34590.1"
/db_xref="GI:13360627"
/translation="MVFAKSPARSKTAGTTTCALKESDMAIAASYTMHLYCDCLQCTD
GKYKSPDFGEYIGTSHAGCAKEARKDGRISRDKTRAFAPGHKILRSNKEG"
/complement(112609. .113076)
/gene="Ecs1168"
/complement(112609. .113076)
/gene="Ecs1168"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881600|dbj|BAA84291.1| percent
identity 100 in 155 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34591.1"
/db_xref="GI:13360628"
/translation="MLRPTLIAMHRLTKSLWRFTSGFGNKVPSAPLAPASKEVSNA
CGMYRLVTVSDFLGCANCCKEHHPDTPKAPPYCSGTTSSKEMQESKTMKRRKAKIL
LVRNAPGVQWVRLSNRRRMGLMKYIGMMDCGFCCKPQSAQNWKHNLRTKGE"
/complement(112949. .113722)
/gene="Ecs1169"
/complement(112949. .113722)
/gene="Ecs1169"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585386|gb|AAD25414.1|AF125520_9
percent identity 100 in 257 aa, similar to hypothetical
proteins e.g. [Bacteriophage 933W]
gi|4585455|gb|AAD25483.1|AF125520_78 percent identity 95
in"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34592.1"
/db_xref="GI:13360629"
/translation="MSEINYQALREKAERKATGSIYVGHTSVNHQNLGTGVFCQKWK
GEPGVIAECHVNCILTESDQVAEANAFTAEANPATVVALLDQERNQOYIKRRDOEN
EETALTVGKLRVELEAENNLIDSECHVALEALRDQALLSEAKRNKAKLOSENAY
IRNRYKELDLICKNILVQQAALIEWQATGDAKSLAWIYNTLFGGPELPDESEKDAQ
AYFNRYAPIDEKLMALHMKWFQESAEARAAGIRIKRGE"
/complement(113719. .113940)
/gene="Ecs1170"
/complement(113719. .113940)
/gene="Ecs1170"
/notes="unknown, similar to C4-type zinc finger proteins
(Trar family) e.g. gi|4585456|gb|AAD25484.1|AF125520_79
percent identity 79 in 73 aa"
/codon_start=1
```

```

/transl_table=11
/evidence-not_experimental
/product="C4-type zinc finger protein (Trar family)"
/protein_id="BAB34593.1"
/db_xref="GI:13360630"
RR1AVGCRTCASCQDELEISKQSK"
complement(114039..114254)
/gene="ECS1171"
CDS
complement(114039..114254)
/gene="ECS1171"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] but different start, also similar to
orf61 [Bacteriophage lambda] gi|508993|gb|AAA96566.1|
percent identity 93 in 46 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34594.1"
/db_xref="GI:13360631"
/transl_table="MTPQENALRSIARQANSEIKKARQOFDPKNVDDICRSVLKKHR
ETVLMGFTTHLSLAIGMLNGVFKER"
complement(114331..114522)
/gene="ECS1172"
CDS
complement(114331..114522)
/gene="ECS1172"
[note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881603|dbj|BAA84294.1| percent
identity 100 in 63 aa, also similar to orf63
[Bacteriophage lambda] gi|508994|gb|AAA96567.1| percent
identity 90 in 61 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34595.1"
/db_xref="GI:13360632"
/transl_table="MKKASPELRTSIEMAHSLAQIGRVFPIPVETDEEFHTLATSL
SQLEMMAAKAEANERDPA"
complement(114495..114677)
/gene="ECS1173"
CDS
complement(114495..114677)
/gene="ECS1173"
[note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881604|dbj|BAA84295.1| but
different start, also similar to orf60a [Bacteriophage
lambda] gi|508995|gb|AAA96568.1| percent identity 96 in 60
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34596.1"
/db_xref="GI:13360633"
/transl_table="MKHPHDNIRVGAI TFVSYVTKRGWVFPGLSVIRNPLKAORLAE
INNKRAVCTKHLPLS"
complement(114674..115354)
/gene="ECS1174"
CDS
complement(114674..115354)
/gene="ECS1174"
[note="exonuclease, similar to exonuclease [Bacteriophage
lambda] gi|2981722|pdb|1AVQ|A percent identity 98 in 226
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="exonuclease"
/protein_id="BAB34597.1"
/db_xref="GI:13360634"
/transl_table="MTPDIILQGTIDVRAVEQGDAAWHLRLGVTASEVHNIAKP
RSGKWPDKMSYFHTLLAEVCTGVAPEVNAKALAWKOYENDATLFFETSGVNVIE
SPIIYRDESMRTACSPDGLCSGNGLELKCPFTSRDFMKFRLLGGFAIKSAYMAQVQY

```

```

SMVYTRKDAMYFANYDPRMKREGHYVVVERDEKYMASFDEWVPEFTEKMDALAEIG
FVFGQMR"
complement(115351..116136)
/gene="ECS1175"
CDS
complement(115351..116136)
/gene="ECS1175"
[note="recombination protein Bet, identical to Bet
[Bacteriophage VT2-Sa] gi|5881606|dbj|BAA84297.1| percent
identity 100 in 261 aa, also similar to Bet [Bacteriophage
lambda] gi|137511|sp|P03698|VBET_LAMB percent identity 99
in 261 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="recombination protein Bet"
/protein_id="BAB34598.1"
/db_xref="GI:13360635"
/transl_table="MSTALATLAKGLAERVGMSVDQELITLRLQRTAFKGDASDAQF
IALLIVANQGLNPWTKEIYAFDPKQNGIVPVGVGWSRIINENQOFGDMDFQDNE
SCTCRIYKDRNHPICVTEWMDCECRPFRTREGREITGPWQSHPKRLRHKAMIQCA
RLAFAGIYDKDEAERIVENTAYTAERQPERDITPVNDETMOEINTLLIATDKTWD
DLLPLCSOIFPRDIRASSELTAQAEAVKVLGFLQKQKASEQKVA"
complement(116142..116558)
/gene="ECS1176"
CDS
complement(116142..116558)
/gene="ECS1176"
[note="host-nuclease inhibitor protein Gam, similar to Gam
proteins e.g. [Bacteriophage lambda]
gi|138128|sp|P03702|VGAM_LAMB percent identity 97 in 138
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="host-nuclease inhibitor protein Gam"
/protein_id="BAB34599.1"
/db_xref="GI:13360636"
/transl_table="MDINTETEIKQKSLTPFPVLISPAFRGRYFHSYFRSSAMNAY
YIADLEAOSQTHRYQOIAEREKEALADDMGKGLPQHLFESLCIDHLQRHGASKRAI
TRAFDDVLFQERMAHRIYVETIAHQVDIDSEV"
complement(116513..116782)
/gene="ECS1177"
CDS
complement(116513..116782)
/gene="ECS1177"
[note="Kil protein, identical to kil [Bacteriophage VT2-Sa]
gi|5881608|dbj|BAA84299.1|, similar to kil proteins e.g.
[Bacteriophage lambda] gi|138622|sp|P03758|VKIL_LAMB
percent identity 98 in 89 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="Kil protein"
/protein_id="BAB34600.1"
/db_xref="GI:13360637"
/transl_table="MPLOGLLLAALPNLYLNESPVYVTDGNALSTYLISQESQKMD
QILMAQITATFATFIDGDMKFEADVAKKWLILKLRSSKSIH"
complement(116625..116789)
/gene="ECS1178"
CDS
complement(116625..116789)
/gene="ECS1178"
[note="regulatory protein cIII (antitermination), identical
to cIII [Bacteriophage lambda]
gi|133366|sp|P03044|RPC3_LAMB percent identity 100 in 54
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="regulatory protein cIII"
/protein_id="BAB34601.1"
/db_xref="GI:13360638"
/transl_table="MQYAIAGWPVAGCPSESLERITRKLDRGWKRLDILNQPGVPK
NGSNVYGYDP"
complement(116862..117230)
/gene="ECS1179"

```


CDS

```
complement(116862. .117230)
/genes="ECs1179"
/notes="single strand binding protein Eal0, identical to
Eal0 [Bacteriophage VT2-Sa] gi|5881610|dbj|BAA84301.1|
percent identity 100 in 122 aa, similar to Eal0
[Bacteriophage lambda] gi|137630|sp|P03757|VEL0_LAMB0
percent identity 99 in 122 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="single strand binding protein Eal0"
/protein_id="BAB34602.1"
/db_xref="GI:13360639"
/translation="NSNIKKYIIIDYDWKASIEFIEDHVDVTEEKLHQINNFWSDEYR
LNKHSGLNAVLIAMLAQHALIIASSDLNAYGVCFDWNQNGQEGWPPMDGSEGR
ITDIDTSGIFDSDDMTIRAA"
```

gene

```
complement(117413. .117664)
/genes="ECs1180"
```

CDS

```
complement(117413. .117664)
/genes="ECs1180"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881612|dbj|BAA84303.1| percent
identity 100 in 83 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34603.1"
/db_xref="GI:13360640"
/translation="MEFHSAICDFRANSVKQPIAVLFKTMGAWAVLCFAADTD
ARMAIQGEMDPTNDEFIYGAPSNYLLDTCNIYNKAA"
```

gene

```
/genes="ECs1181"
/genes="ECs1181"
```

CDS

```
complement(117723. .117995)
/genes="ECs1181"
/notes="probable anti-termination protein N, identical to N
protein [Bacteriophage VT2-Sa] gi|5881613|dbj|BAA84304.1|
but different start, similar to N proteins e.g.
[Bacteriophage 933W] gi|4585397|gb|AAD25425.1|AF125520_20
percent identity 42 in 90 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative anti-termination protein N"
/protein_id="BAB34604.1"
/db_xref="GI:13360641"
/translation="MSRKTEFGTAASRRAPRANLQOEAISSDKLHRPTFSRVLQ
CKRKPAARAEVITLTLRKVEGSTCLPNVALYAAGYRKSQLTAR"
```

gene

```
complement(117973. .118155)
/genes="ECs1182"
```

CDS

```
complement(117973. .118155)
/genes="ECs1182"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34605.1"
/db_xref="GI:13360642"
/translation="MEQTGRLFKQRLSTTWLKSQITOPHKLWDAMPKQPSOEELRDC
IAKVYSGGIYVOKNRI"
118452. .118613
/genes="ECs1183"
118452. .118613
/genes="ECs1183"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34606.1"
/db_xref="GI:13360643"
/translation="MLTHMKISMYPYFPIGHIHNFISIELVIOIVCRHLCFLILAAALQ
```

gene

```
118452. .118613
/genes="ECs1183"
```

CDS

```
118452. .118613
/genes="ECs1183"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34606.1"
/db_xref="GI:13360643"
/translation="MLTHMKISMYPYFPIGHIHNFISIELVIOIVCRHLCFLILAAALQ
```

gene
CDS

```
PLEKMYQTR"
complement(118724. .119245)
/genes="ECs1184"
complement(118724. .119245)
/genes="ECs1184"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881614|dbj|BAA84305.1| percent
identity 100 in 173 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34607.1"
/db_xref="GI:13360644"
/translation="MAKSNVSVQAFKDFLEELMSLNIMKEATARNLKNSSARLLTVVQ
EEMGDVTDVNELAERYINATEPKPSSSITAYKRMESATKKVAFQSGGEIPLYT
PIDKESSEKDLTGPTKVEGRKANALHTDLPVLRPESGVTVTIKIPNDITNEAE
RISSILKVIYRFQ"
complement(119747. .120400)
/genes="ECs1185"
complement(119747. .120400)
/genes="ECs1185"
/notes="probable CI repressor protein, similar to CI
[Bacteriophage lambda] gi|133353|sp|P03034|RPC1_LAMB0
percent identity 70 in 208 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative CI repressor protein"
/protein_id="BAB34608.1"
/db_xref="GI:13360645"
/translation="MKWYELARSRMKELGITQKLAELMTQGGIGHWLRGSRHPSL
SDIGVVFYLGIDINISFNHDGTFSPVGEYSAPVKQYEPVPSHVQAGMFSPELRTF
TKGDAERLYSTTKASDSAPFLEVEGNSMTAPTGSKPFDGMLIIVDEQAVPEQDF
CIARLGGDEFTRKFLIRDSQVFLQPLNQYPMIPCNESCSVVGVKIAGQWPEETFG"
120518. .120733
/genes="ECs1186"
120518. .120733
/genes="ECs1186"
/notes="probable regulatory protein, identical to
hypothetical protein [Bacteriophage VT2-Sa]
gi|5881616|dbj|BAA84307.1|, similar to c2 [Bacteriophage L]
gi|1469215|emb|CAA63999.1| percent identity 42 in 49 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative regulatory protein"
/protein_id="BAB34609.1"
/db_xref="GI:13360646"
/translation="MSNLRKYRESLNISQTTLAKAVGCTQGAIGHWESGRRRFPDLKTC
RALVACLNLKGAKVSLDDVFPPEHKAA"
120875. .121171
/genes="ECs1187"
120875. .121171
/genes="ECs1187"
/notes="regulatory protein CII, identical to CII protein
[Bacteriophage VT2-Sa] gi|5881617|dbj|BAA84308.1| percent
identity 100 in 98 aa, similar to CII proteins e.g.
[Enterobacteria phage HK022] gi|631957|pir||S42398 percent
identity 96 in 98 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="regulatory protein CII"
/protein_id="BAB34610.1"
/db_xref="GI:13360647"
/translation="MEQTSYKLSQREIDRAETDLLINLTQRLAKMIGCHESKI
SRTDWRFIASVLCFAFGMSADISPIRAFAYALDEITKKSPAATEDFKIDMQF"
```

gene
CDS

```
121204. .121350
/genes="ECs1188"
121204. .121350
/genes="ECs1188"
/notes="unknown, identical to hypothetical protein
```

```

[Enterobacteria phage HK022] gi|632160|piri||S42399 percent
identity 100 in 48 aa, also similar to orf48
[Bacteriophage P22] gi|871503|emb|CAA55155.1| percent
identity 85 in 48 aa
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34611.1"
/db_xref="GI:13360648"
/translation="MTKQLSPYQDKTHKHILRDLRFSLSPKQPGFRFAELEKVKLMQKE
KGHE"
gene
121343..122242
/genes="ECs1189"
CDS
121343..122242
/genes="ECs1189"
/notes="replication protein O, similar to O proteins e.g. O
[Enterobacteria phage HK022] gi|407289|gb|AAB60272.1|
percent identity 98 in 299 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="replication protein O"
/protein_id="BAB34612.1"
/db_xref="GI:13360649"
/translation="MSNISNLAEARARLRQQHQSGKGAYALLHKKIMDVPPYKDAE
AHLVHLILKAKHTPEYMTDAGEILVGRKLLGGRNSLAFETGLKPRQVLLRRF
KILGMDVSHGKFSVSEKYDYQSNFVADYQIITTSKPAIMPASNTVPADYQQ
ITTDREYNNIISNTDVLSEATADKKSKKPSVSCQVDVADYHEILLPEAPKIRALNDK
RKNOIRTEWRKAGVITRLDGHGFTMQDWRNYLSYVGENCRMFEERPQHQRCTVWHK
KGFDELLDNTYLVKREGEHDDR"
gene
122217..123668
/genes="ECs1190"
CDS
122217..123668
/genes="ECs1190"
/notes="replication protein P (putative replicative DNA
helicase), similar to P proteins e.g. [Enterobacteria phage
HK022] gi|6863143|gb|AAF30384.1|AF069308.32 percent
identity 99 in 478 aa, also similar to replicative DNA
helicases e.g. DnaB [Escherichia coli]"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="replication protein P"
/protein_id="BAB34613.1"
/db_xref="GI:13360650"
/translation="NRVNTMTDNFYAPPHSIEAEQAVIGGLLLDDSSSRVOKVLAML
KPSFYSRPHKIIFEIITRMHREQKPVDTGLTFLDELRKSLTASVGGFAYIAEIAKNT
PSAANIYAVAMOVRETAMERYAINRMTEATELLYSRNGMTATQKYEAOIAITQLTDH
AKTSRGLRSRGEVWEDVSLERFDSGQRGSGTGISLDRMLSPKGLVKGLSF
VIGARPKMKTLLYSOMAINCAVHEKRPALMFSLPMGDOILEKLVGQKSGVNPNIIFY
LPATNDADGYQGDYDGFNFRAITANRLSEIDLILYIDTPGLSLAQIVSESRRIKRE
KCGVMILVDYLLTMTAEKADRNLDLAYMITKGLKNLAKELDCVVLVLTOLNRALESR
TNKRPLPSDRSGTOIEODCYWVGIIHREGAFDDSVPPGETELILRLNRHGNTGTVYC
IQANGAIVTDQGSAAEMRREREREQSKKKGF"
gene
123668..123937
/genes="ECs1191"
CDS
123668..123937
/genes="ECs1191"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881620|dbj|BAA84311.1| percent
identity 100 in 89 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34614.1"
/db_xref="GI:13360651"
/translation="MNKKQALILEKAWDAQISYALKEQALPTIOTKSKTARQLCDGGF
LNEIETROMVTFKGYEINHHGIAAYCSHLPPDDVDIDEMEREMKQ"
gene
124008..124286
/genes="ECs1192"
CDS
124008..124286

```

```

/genes="ECs1192"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4499788|emb|CAB39287.1| percent
identity 100 in 92 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34615.1"
/db_xref="GI:13360652"
/translation="MNDYSRQFENWWSKDSQFTGDDDELKEFAWVWQASRSATLELDI
DWPESENDPFWKDEEGAYAMGYEDGRDKTIVAMKAIARAAGIKEKNEFD"
gene
124419..124634
/genes="ECs1193"
CDS
124419..124634
/genes="ECs1193"
/notes="unknown, similar to hypothetical
protein[Bacteriophage 933W] gi|4499789|emb|CAB39288.1|
percent identity 97 in 71 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34616.1"
/db_xref="GI:13360653"
/translation="MSTIAELVRANFREELVRWYRVRSSSSSLPLDELYEHSPAARRYP
RDRVLRLFLKLNNEFORNRIIRSLDK"
gene
124645..124881
/genes="ECs1194"
CDS
124645..124881
/genes="ECs1194"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881622|dbj|BAA84313.1| but
different start, also similar to hypothetical proteins
[Bacteriophage 933W] gi|4499790|emb|CAB39289.1| percent
identity 85 in 78 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34617.1"
/db_xref="GI:13360654"
/translation="MSDLSLTPKLEKPCPGGNARLWVEAGINIDVWGYAECDLCEA
RGAMAPSVAAAEKWNRRAGDEANLSASQRSNOK"
gene
124838..125284
/genes="ECs1195"
CDS
124838..125284
/genes="ECs1195"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881623|dbj|BAA84314.1| but
different start, also similar to hypothetical proteins e.g.
NinB protein [Bacteriophage 21] percent identity 43 in 147
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34618.1"
/db_xref="GI:13360655"
/translation="MKQTELLRNEATRNNADILSLPIDDKSPHEVHVHVEPKRSRAQ
NORMWMLNDVSRQVLWHGRLAPEDWKDLFTALWLKTRKLEQSRVPGIDGVVMLGV
RYSKMRKASMTLIEIMFWFGSERNVRWSDSRREYSQRKGRAA"
gene
125281..125808
/genes="ECs1196"
CDS
125281..125808
/genes="ECs1196"
/notes="probable DNA methylase, identical to hypothetical
protein [Bacteriophage VT2-Sa] gi|5881624|dbj|BAA84315.1|
percent identity 100 in 175 aa, similar to hypothetical
protein Gp62 [Bacteriophage HK97]
gi|6901634|gb|AAF31137.1| percent identity 98 in 175 aa"
/codon_start=1
/transl_table=11

```

```
/evidence=not_experimental
/product="putative DNA methylase"
/protein_id="BAB34619.1"
/db_xref="GI:13360656"
/translation="MTIKNSTPAHDKDCWQTPFLWLFDAIDIEFGFWLDSAAADKNAIC
AHMLTEADALNSEVSHCAINNPYSNIRPWVEKAAEOCLOQRTQVYVLMVPEMSV
GWFSALESDEVRIITDGRINFIEPSTCLEKAGKSGMLLIWRPFISPRNMFITVS
KAALMAIGOVVRAA"
gene
125805..125987
CDS
125805..125987
/gene="ECs1197"
/gene="ECs1197"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881625|dbj|BAAB4316.1| percent
identity 100 in 60 aa, similar to hypothetical proteins
e.g. NINE protein [Bacteriophage 21]
gi|4539480|emb|CAB39989.1| percent identity 98 in 60 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="NINE protein"
/protein_id="BAB34620.1"
/db_xref="GI:13360657"
/translation="MRORRSITDIICENCKYLPYKRSNKKRPDKESDVKTFTYTA
HLMIDRLRHRARKTR"
gene
126092..126265
CDS
126092..126265
/gene="ECs1198"
/gene="ECs1198"
/notes="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] gi|5881626|dbj|BAAB4317.1| percent
identity 100 in 57 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34621.1"
/db_xref="GI:13360658"
/translation="MFFIVVCTRTSELWVWVGRRDRPRSRLRQPCTVHHQTIGVV
GGELLNRHORASL"
gene
126262..126996
CDS
126262..126996
/gene="ECs1199"
/gene="ECs1199"
/notes="probable antirepressor protein, identical to
hypothetical protein [Bacteriophage VT2-Sa]
gi|5881627|dbj|BAAB4318.1| percent identity 100 in 244 aa,
C-terminal part is similar to C-terminal part of
antirepressor protein Ant [Bacteriophage P22] gi|13184"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative antirepressor protein"
/protein_id="BAB34622.1"
/db_xref="GI:13360659"
/translation="MTQISVETLSPIITHNQIPVITTELLAOLYGTPEVPIRONHHEN
KVREVEGHFKVGNLDKELRVALNSQNLVTSNQNPQSLRGLQISPKARSLI
LWTEGARHAKMLETDAQWDFEKLDCYSQCEKNTGQKELNGLSARETDSLW
LWDYANRSCALFREDYPAKLIQSGYSGICHIDGYEFYSIIIGRARGVNLINHTROIDIY
EPDGTPTNLLAWERLKNKPELSLHY"
gene
127071..127793
CDS
127071..127793
/gene="ECs1200"
/gene="ECs1200"
/notes="DNA-binding protein, identical to Roi [Bacteriophage
VT2-Sa] gi|5881628|dbj|BAAB4319.1| but different
start, similar to Roi proteins e.g. [Enterobacteria phage
HK022] gi|1197729|gb|AAC48863.1| percent identity 82 in
242 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="DNA-binding protein"
```

```
/protein_id="BAB34623.1"
/db_xref="GI:13360660"
/translation="MNELINSNAIKMTSTIEAELVSGRSHDKVKQSTLERLAVRGVIRNP
PVMVEKINNLGLRGVEAYVEGEGKRDSTIIVVAQLSPEFTARLVDRMRLEGATA
KTPQFSEALRLAALDEDOKALEKQALAAAPKVPADRVGEASGLIGNAKVVGIG
PNKLFANMRDHKILLIAGSERNVPMQEYMDRGYFTVKETAVENTNHGIIQISFTTKITGR
GOOMLTRKLLDNGMLUKVTREAA"
gene
127793..128398
CDS
127793..128398
/gene="ECs1201"
/gene="ECs1201"
/notes="unknown, identical to hypothetical protein orf15
[Bacteriophage 93W] gi|4499798|emb|CAB39297.1| percent
identity 100 in 201 aa, similar to hypothetical proteins
e.g. Ning protein [Bacteriophage 21]
gi|4539482|emb|CAB39991.1| percent identity 94 in 20"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein Ning"
/protein_id="BAB34624.1"
/db_xref="GI:13360661"
/translation="NAKPARRKCKICKWFHPAFSNQWCCPEHGTQLALERSKERE
KAEKAAERARRREORQDKLIRKLALAPRSTWIKQAQAVNAFIRERDRLPCISC
GTLTSAQWDAGHYRTAAAPQLRFDERNIIHKQCVNQHKSGLNLYPYRVELINRIQGE
AYDEIESNNHRHWTVEECAIKAKYQQKLIKDLNRSRSEA"
gene
128395..128589
CDS
128395..128589
/gene="ECs1202"
/gene="ECs1202"
/notes="unknown, identical to hypothetical protein orf16
[Bacteriophage 93W] gi|4499799|emb|CAB39298.1| percent
identity 100 in 64 aa, similar to hypothetical proteins
e.g. Ning68 [Bacteriophage lambda]
gi|9626304|ref|NP_040640.1| percent identity 80 in 60 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34625.1"
/db_xref="GI:13360662"
/translation="MTFTVKTIPDMLVEAYENQTEVARILNCSRNVTYRKYTGDKGKR
HAIVNGVLMVHRGWGKDTDA"
gene
128582..129016
CDS
128582..129016
/gene="ECs1203"
/gene="ECs1203"
/notes="antitermination protein Q, identical to
antitermination protein Q [Bacteriophage 93W]
gi|4585416|gb|AAD25444.1|AF125520.39 but different
start, similar to antitermination Q proteins e.g.
[Bacteriophage H-19B] gi|2668768|gb|AAD04655.1| percent
identity"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="antitermination protein Q"
/protein_id="BAB34626.1"
/db_xref="GI:13360663"
/translation="MRDIRQVLRGMAWANNYEDVTWSPIAAGFKGLIPEKVKSRQP
CCDDAMVICGICARLYRNNRDLHLLVDYVYVLTGTFMALARKHGCSDTICIGRLHKA
EGIVGEGMLMLGVRLMDRYVERELPGGRTSVFYQRKNSLRS"
gene
129265..129417
CDS
129265..129417
/gene="ECs1204"
/gene="ECs1204"
/notes="unknown, similar to part of hypothetical protein
[Bacteriophage P27] gi|8346570|emb|CAB93763.1| percent
identity 89 in 37 aa, TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
```

```

/protein_id="BAB34627.1"
/db_xref="GI:13360664"
/translation="MPASPSGWLFFVYRSVKAAAMALGRRRAIGVLESGRFEQTREVQNV
VSQNG"
129458..129533
/gene="ile22"
/notes="anticodon: CAU, Cove score 88.46"
/product="tRNA-Ile"
129458..129533
/gene="ile22"
129543..129619
/gene="argN2"
/notes="anticodon: UCG, Cove score 61.15"
/product="tRNA-Arg"
129543..129619
/gene="argN2"
129633..129709
/gene="argO2"
/notes="anticodon: UCU, Cove score 83.94"
/product="tRNA-Arg"
129633..129709
/gene="argO2"
129800..130759
/gene="ECs1205"
129800..130759
/gene="ECs1205"
/notes="Shiga toxin 2 subunit A, identical to
gi|1351074|sp|P09385|SLTA_BP933"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Shiga toxin 2 subunit A"
/protein_id="BAB34628.1"
/db_xref="GI:13360665"
/translation="MKCILFKVWLCLLGFSSVSYSREFTIDFSTQOSVSVSLNSIRT
EISTPLEHSQCTTSVSNHTPGSYFAVDIRGLDYOAREDLRLIIEQNNLYVAG
FYNTAINTFYRDSFTHSVPGVTVMSTTDSSTYTLQORVAALERSGMOISRSLVSS
YIALMEFGSNTWRDASRAVLREVTVAALRFRIQREFRALSETAFVITMTGCDV
DLTLNWRGRISNVLPEYRGEDGVVRGRISFNNSAILGTVAIVILNCHHGAGRSVRVNE
ESQPCQITGDRPVIKINNTLWESNTAAFLNRKSFQFLYTCK"
130771..131040
/gene="ECs1206"
130771..131040
/gene="ECs1206"
/notes="Shiga toxin 2 subunit B, identical to
gi|134538|sp|P09386|SLTB_BP933"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Shiga toxin 2 subunit B"
/protein_id="BAB34629.1"
/db_xref="GI:13360666"
/translation="MKKMFMAVLFFALASVNAAMADCAKGKIEFSKYNNEDDTFTVRKVDG
KEYWTSRNLOPLQSQALTGMTVTIKSSTCSGSGFAEVQFNND"
131527..133431
/gene="ECs1207"
131527..133431
/gene="ECs1207"
/notes="unknown, identical to N-terminal part of
hypothetical protein [Bacteriophage 933W]
gi|4585419|gb|AAD25447.1|AF125520.42 percent identity 100
in 557 aa, similar to N-terminal parts of hypothetical
proteins e.g. Yjhs [Shigella dysenteriae] gi|6759965|gb|
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34630.1"
/db_xref="GI:13360667"
/translation="MAFKHYDVRRAASPSDLAKRITQKKEGWQPGYSALISTAGYGA
BFIPQVPEVTSISSEPEGNRPFTTSAPVAPEYVYVIALAGOSNGMSYEGGLPLPDTF
SDPPRIKOLARSTVTPGGAVCKYNDIIPADHCILHDVQDMRLNHPKADLSKQGYGT
VGOGLHIAKLLPFIIPANAGILLVPCCRGSAFTTGADTYSADASGASENSTRWGVDK

```

```

PLYKDLIGRTKAALKKNPKNVLFVAVVMQGEFDDGTPVNHAAQFGALVDKFRADLAD
MAQCQVGSAGVPWICGDTTYFMKQKNSTYQTYVGSYKNKTEKNIHFVPFMTDENG
VNYPTNKPEDDPDIPGIGYIGSKWRDSSATWTQDRASHFSSWARRGIISDRLATAIL
RHAGRVNALNAGASSTVEVSPSPSGAEATGVTTLLSYLASESGSLKVOGWSAGGR
AEVVSDAEGTGGRKAVKITKEAGKSSWVLEVAAGNALLQKGGQIRCFKRVSGALAA
QYVMATYVPVSSILPOGVALTGDGNNLLAIFYITDAKDLNVMYHNKAVNRPGPGEC
FICELRLPDHFRWRKHNKLELLPEYGPAPPAIVDCYTSPTTLVMPVPLLHGFSSSY
GVLPPLCKDH"
complement(133196..134505)
/notes="IS629-02"
134128..134128
/gene="ECs1208"
complement(133238..134128)
/gene="ECs1208"
/notes="transposase (insertion sequence IS629), identical to
gi|7443862|pir||T00240"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase OrfB protein of insertion
sequence IS629"
/protein_id="BAB34631.1"
/db_xref="GI:13360668"
/translation="MPLLDKLRQYGVGPVCSSELHIAPSTYHCOQORHHHPDKRSAR
AQHDDWLKREIORVYDENHQVGVGRVWRQLLREGIRVARCTVARLMAVMGLAGVLRG
KKVYTTISRKAAGADRVNRQFVAERPDQLMWADFTVSTWQGVYVAFITDVFAGYI
VGMRVSSMETTFVLDALQALWARRPSTIIHSDKGSQVSVSLAYTERLKEAGLLAST
GSTGDYDNAAEISINGLYKAEVIHRKSKNRAEVELATLTWVDWYNRRLLRLGLHT
PAAEAKYVASIGNDLAA"
complement(134125..134451)
/gene="ECs1209"
complement(134125..134451)
/gene="ECs1209"
/notes="transposase (insertion sequence IS629), identical to
gi|7444868|pir||T00241 percent identity 100 in 108 aa"
/codon_start=1
/evidence=not_experimental
/product="putative transposase OrfA protein of insertion
sequence IS629"
/protein_id="BAB34632.1"
/db_xref="GI:13360669"
/translation="MTKNTRFSPVQRQRAIRMVLESQDEYDSQWAAICSIAPKIGCTP
ETLRVWVROHERDTGGGDDGLTSAERQRLKELERENRELRNSDILRQASVFAKAEF
DRLMKK"
134512..134778
/notes="partial CDS, identical to C-terminal part of
hypothetical protein [Bacteriophage 933W]
gi|4585419|gb|AAD25447.1|AF125520.42 percent identity 100
in 90 aa, also similar to C-terminal parts of hypothetical
proteins e.g. Yjhs [Shigella dysenteriae]
gi|6759965|gb|AAF28123.1|AF153317.19 percent identity 53
in 83 aa, disrupted by IS insertion"
134913..135092
/gene="ECs1210"
134913..135092
/gene="ECs1210"
/notes="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4499806|emb|CAB39305.1| percent
identity 100 in 59 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34633.1"
/db_xref="GI:13360670"
/translation="MTFLNQLMLYFCTVVCVLYLLSGGYRAMRDFWRROIDKRAAEKI
SASQSGAGSKPEEPLI"
134959..135405
/gene="ECs1211"
134959..135405
/gene="ECs1211"
/notes="unknown, identical to hypothetical protein

```

```

[Bacteriophage 933W] gi|4585420|gb|AAD25448.1|AF125520_43
percent identity 100 in 148 aa
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34634.1"
/db_xref="GI:13360671"
/translacion="MCCISFRVGTGCVTSGADRLTKGPIRKSPVQPEANPKSRSF
SGNFLNHTFRRENPMSEITSLVTAEVKDVLRSEVSALKOKLRHLEARLDAEYDA
ILDELGVQAPPTPEAGDTTAESEGVQPEFVADATEPQESVNNL"
135482..135697
/gene="ECs1212"
135482..135697
/gene="ECs1212"
/notes="probable holin protein, identical to protein S
[Bacteriophage vT2-Sa] gi|5881636|dbj|BAA84327.1|similar
to putative holin proteins e.g. [Shigella dysenteriae]
gi|16759967|gb|AAE28125.1|AF153317_21 percent identity 95
in 71 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative holin protein"
/protein_id="BAB34635.1"
/db_xref="GI:13360672"
/translacion="MYOMEXITTVGSVYTSVAVGTGYWFLQLLDVRVSPSQAAIGVLGS
LFLGLLTLYNLNLYFKIKEDRRKAARGE"
135702..136235
/gene="ECs1213"
135702..136235
/gene="ECs1213"
/notes="endolysin, identical to putative endolysin
[Bacteriophage 933W] gi|4585422|gb|RAD25450.1|AF125520_45
percent identity 100 in 177 aa, similar to putative
endolysins e.g. [Bacteriophage H-19B]
gi|1435686|gb|AAD17382.1| percent identity 93 in 177 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative endolysin"
/protein_id="BAB34636.1"
/db_xref="GI:13360673"
/translacion="MSRKLRGLSAVLAIAAGASAPEILDQFLDEKGNHTTAYRD
GAGWITICRATGVDGKVPVPGMKLSKECDRVNATERDKALAWVEKNIKVPLTEPOK
AGTASCPYNIQPGKCFPSTFYRRINAGDRKACEAIRWIKDGRDCRIRSNICYQ
VSRDQESALACWGIDR"
136506..137075
/gene="ECs1214"
136506..137075
/gene="ECs1214"
/notes="putative antirepressor protein, identical to
putative antirepressor protein [Bacteriophage 933W]
gi|4585423|gb|AAD25451.1|AF125520_46, similar to
antirepressor protein Ant [Bacteriophage P22]
gi|131843|sp|P03037|IRANT_BPP22 percent identity 49 in 121
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative antirepressor protein"
/protein_id="BAB34637.1"
/db_xref="GI:13360674"
/translacion="MNMMAVPFHGNSLYVYVNHNGEPVPMKPVVAGNGLAQSLAKL
RQRFASITIEVWVAEDGKORNMVSPRLRLAGLQITINPKVKPEIRDKVIRQEEC
DDVLYETWTFGVNPRKMSMEELNQACAMKDRKNIASVFATGLNEWQVRAAHVS
KIRTLVNEANMLIDFVLADTGGKIKTKAD"
137229..137693
/gene="ECs1215"
137229..137693
/gene="ECs1215"
/notes="probable endopeptidase, identical to endopeptidase
[Bacteriophage 933W] gi|4585424|gb|AAD25452.1|AF125520_47
```

```

percent identity 100 in 154 aa, similar to endopeptidases
e.g. Rz [Bacteriophage lambda]
gi|19368|sp|P00726|ENPP_LAMBD percent identity 72 in"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative endopeptidase"
/protein_id="BAB34638.1"
/db_xref="GI:13360675"
/translacion="MNRVLCVVIIVLAVGYCALMLATNHYBDNALTYKAQRKKAREL
EQANATIDMQVRORDVAALDAKISRELADARAENETLRADVAGRKKLRINATCSGT
VREATGTSGVDNATGPRLADTAERDYFILRELITMQKLEGTKQYINEQCR"
137449..137634
/gene="ECs1216"
137449..137634
/gene="ECs1216"
/notes="probable lipoprotein R21 precursor, identical to
putative R21 protein precursor [Bacteriophage 933W]
gi|4585425|gb|AAD25453.1|AF125520_48 percent identity 100
in 61 aa, similar to lipoprotein R21 precursor
[Bacteriophage lambda] gi|1017781|gb|AAC48862"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative lipoprotein R21 precursor"
/protein_id="BAB34639.1"
/db_xref="GI:13360676"
/translacion="MRELKMKLCVLMPLVVSACGSTPPAPVPCVKPPAPPATWIMQPA
PDWTPNLNGIISSESG"
complement(137725..138018)
/gene="ECs1217"
complement(137725..138018)
/gene="ECs1217"
/notes="Bor protein precursor, identical to [Bacteriophage
933W] gi|4585426|gb|AAD25454.1|AF125520_49 percent
identity 100 in 97 aa, similar to Bor protein precursor
[Bacteriophage lambda] gi|137520|sp|P26814|VBOR_LAMBD
percent identity 96 in 97 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Bor protein precursor"
/protein_id="BAB34640.1"
/db_xref="GI:13360677"
/translacion="MKKMLATALALITGCAQOTFTVONKQTAVAPKETITTHFFVFS
GIGGKTVDAKICGCTENVVKTETQTFVNGLLGFTLGIYTPLEARVYCSQ"
138138..138371
/gene="ECs1218"
138138..138371
/gene="ECs1218"
/notes="unknown, similar to hypothetical protein
[Bacteriophage vT2-Sa] gi|5881640|dbj|BAA84331.1| percent
identity 85 in 75 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34641.1"
/db_xref="GI:13360678"
/translacion="MNRHCIPGLTMTALTGTGNFSAGVCGDNQKRCRTPGFLIFHEMGAI
SREAAQVPRWMKPKDTKALTCKYRSNMVLLCEI"
138427..139233
/gene="ECs1219"
138427..139233
/gene="ECs1219"
/notes="small subunit terminase, identical to
putative small subunit terminase [Bacteriophage 933W]
gi|4585427|gb|AAD25455.1|AF125520_50 percent identity 100
in 268 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative small subunit terminase"
```

/protein_id="BAB34642.1"
/db_xref="GI:13360679"
/translation="MAKLDWMKKLEQAFRRREHAETGITLLDWCRRKKKINYNTATRIKM
KPSIDDKHTDHEIDHDIIDEPCNDAGSDGDERCAKSNKNCANSATRIKRSRL
PPSNAFQSRNTHAVRHRYGAKYLEADNMDSDMVLFDLVTFRARALSVTALKGM
FADLEEDATDVETRVLYDKILKABQALDRNARTIESTERSILLTDLVLAETAPKLADR
ERINAARDKURAEITDILTNRQGVVTPVSDIVSSLHEMSNSGRLLDDIPEE"
139214..140920
/gene="ECs1220"
139214..140920
/gene="ECs1220"
/note="probable terminase large subunit, identical to
putative terminase large subunit [Bacteriophage 933W]
gi|4585428|gb|AAD25456.1|AF125520_51 percent identity 100
in 568 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative terminase large subunit"
/protein_id="BAB34643.1"
/db_xref="GI:13360680"
/translation="MTERKNEPRDEPSEMTEAQRRLFIMTKLSNPWRLNHLKYKION
EKGLVTFMRPAQROLFRRSMHNKNIILKARQLGFSTAIDYLLDQALFIPHLKCGIV
AODKQAASEIFRTKIAVPFDHLPDMLRASITIVERRSGASGGYILFCHGSGIOVATSE
RSQTVQRHLHISEGKICAKYPAKAKELRTGTNLAVSECIIFDESTAEQGGDFYEMS
NRAQEITASGLLLTAQDIKFHFYAWQDPKYSARVPESGLKLSREKMTYFSAVEKAMN
ITLDEQKQYINKTEQREEMKQEFPTQEAFLTSGRRVFSAESTLQAESFCSPPM
IVDIEPTVAKTKAQALREGNKNELQRLTMNLYLWELPDPDEYVCGADTAEGLEH
GDRSLSDVKSNGSEQVAHFWGLHDAELFAHLISQVCRMYNNAPVGPERNHGHAVIL
KLRELYPTRIYNEOHLDOAYDDDTPLRGLWTTTQSKPVLTEGKTLILNNGISGIRWS
GLTSEMPTYVDKAGSMNAQGGFDDQLMSYVIAQEMRAPVRVKOKTDKRRTHWM
AH"
140920..143064
/gene="ECs1221"
140920..143064
/gene="ECs1221"
/note="probable portal protein, identical to putative
portal protein [Bacteriophage 933W]
gi|4585429|gb|AAD25457.1|AF125520_52 percent identity 100
in 714 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative portal protein"
/protein_id="BAB34644.1"
/db_xref="GI:13360681"
/translation="MKNETMTMATKNDNGATPRFSQLOALCSDIDSQPKWRDAANK
ACAYYDQQLPPEVLQVLDKRGOPMTIHNLIAPTVDGLGNEAKTRIDLVMSEDPDD
ETEKLABAINAEFADACRLGNMKNRSDAYAEQIKAGLSWVEVRNRSDPPEFKVST
VSRNEFWLSREADLSDCRLMRRMMDTDEAKATFPGMAQVYIDYAIIDWRGFVDT
TVTEGQSPILMSAWEYQSWDRQONLQERRRVLQVYVYRTFERLPVIELSNRVT
VAFDKNLMOAVASGRVQVGVRSVRIEAWFVGFHIVDRPCSAPOGMFLVPFW
GYRDKTGEPYILSRAIPAQDEVNFRRIKLTWLQAKRVIMDEATQISDNDIMEOI
ERPDGIILKLPVRKNQKSVADVFVEQDFQVASQQFVQWSEKLIQDTMGVYSAFILG
QDSGATSGVAISNLVEQGAATTLAEINDNYQFACQVGRLLLYLLDLDLKRNRHAVVI
NRDQRQRTIYNLAEGDNGELNDISRLNTHIALAPVOQTAPAKLAQARMSEVIQV
LPQVQAVLDLWNLIDVPKQEFVERIRAALGTPKSPDEMTPEEEVAQOQALQO
QOAELOREWAGRVAKLEADAAHAAAOORDNASQOREVALTQGRYVVDALNQAHTAE
IITGVONWEOEDVLQOQMLYTLQQRNENSL"
143222..144229
/gene="ECs1222"
143222..144229
/gene="ECs1222"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585430|gb|AAD25458.1|AF125520_53
percent identity 100 in 335 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34645.1"
/db_xref="GI:13360682"
/translation="MDFEFTGEETPEQLEKMLBGLGDVIDIDSHAQDVVVTEDTTEKHAD

BEAQTQTGDNNAVPTPASVEQTQDVKEPEAKGVLTDRGKHVPIYEVLEAERSGKQRA
EQEAALRGQIAEBKRREVELLTSLQHQGMKPTPLPENKISDEQIARIEMYPEIGD
AVASLIIRKNYLOSRVQQAQAEAGGEDLSPLVDAMNAVPLKLTQWESDPDFESVA
VSIQKQNDPAWKDKLTITERFAEVARTQVAFGEVSESSADNKADKTIRKTAEKV
KTAEQEQAVPSPSDLGTTASVGTGDNFERLLGASHSEAEAIMRGMTNAEIDALLEKL
G"
144253..145467
/gene="ECs1223"
144253..145467
/gene="ECs1223"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585431|gb|AAD25459.1|AF125520_54
percent identity 100 in 404 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34646.1"
/db_xref="GI:13360683"
/translation="MTTIVTSAQANKLYQVALFTAAANRRSMVNILTEQOEAPKAVSPD
KKSTQTSAGAPVVRITDLNKGAGDEVTFSTIMHLSKRPMTGDERGEGEDLSHADFE
SLKTNQGRHLVDAGGRMSQORTKFNLAASSATLLGTQFNDLQDOCAIVHLGARGDFV
ADDTILPTAEHPFKIMINDVLPPTHDRHFFEGDATSEFOIEAADIFSIGLVNLSL
FIDEMAHPLQPVRLSGDELHGEDPYVLYVTPPROWNQWYTSISGKDNQMMVRAVNR
KGFNHPFLFKGECAMWRNLIYRKYAGMPLIRFYOGSKVLVSENNLTATTKEVAAATNDR
AMLLGAQALANAYCQAGGHHFNMVEKKTDMDNRTETAIASWINGLKLKTRFPEKSGKMOD
HGVIATVATVKL"
145523..145912
/gene="ECs1224"
145523..145912
/gene="ECs1224"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585432|gb|AAD25460.1|AF125520_55
percent identity 100 in 129 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34647.1"
/db_xref="GI:13360684"
/translation="MAKTIILAPSLSERVYTGTHGNSVAEGVFTVNAEADSVIHLLS
LPVGIRINSIQLVSTGGTGTATVSIKSGEHALIDNSEAVSAKFARYVPEVPTTQRDG
ELVTITKTAATGTGLNVLRYTVVGY"
145962..146423
/gene="ECs1225"
145962..146423
/gene="ECs1225"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585433|gb|AAD25461.1|AF125520_56
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34648.1"
/db_xref="GI:13360685"
/translation="MSBKIAVVIIGPKPVKDDITGSRITLPPRLPEVHVDSAMAWOLL
GFDPVWVRHELDVLLKQOQNEOLROAQOQAEVRLAALAEFNSFVSVNGQVLDLS
KLTSARLATICEAEELDIHKDPKETAFAFRVREAFRRRVAETEQHGGTE"
146407..146970
/gene="ECs1226"
146407..146970
/gene="ECs1226"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585434|gb|AAD25462.1|AF125520_57
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34649.1"
/db_xref="GI:13360686"
/translation="MAELSDFLPYVRRHISGLPINMTDALSMAAVAFSRSLVCCRRE

gene
CDS
gene
CDS
gene
CDS

VTVPVAGKEIVLPYDKDDECECVHIIIRISDNDHELFVGRDVIDSSGRSLRFACSPGEV
SVLYAVAPKAGRSQIPDELLTWPEEVAAGALERLFMTGTGVSWSDDLPLRAQYFVSQFSEG
IRRAYRHTLATSPYSYRNPRVRQRF"

gene
146970..147620
/gene="ECs1227"
146970..147620
/gene="ECs1227"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585435|gb|AAD25463.1|AF125520_58
percent identity 100 in 216 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34650.1"
/db_xref="GI:13360687"
/translation="MTTITELIGRVNTQLVDPMMVRWPLQELCLDYYNDAVRAVILARP
DGASETISCVPGARQVLPDGVILQDLVICLSDGSAVRPLSREVLDQAIPEMPTMKG
IPECFISNDLSRVPFWLFPAPDKREISIDAVWSRIPEAVYVLTDQDDTPVPLEAYVNP
LVEMLFRAFSDAAGAEGLAAQHYQSFEQLGIKQGAADSALYARKRVFNGGV"
147617..149554
/gene="ECs1228"
147617..149554
/gene="ECs1228"
/note="probable tail fiber protein, identical to putative
tail fiber protein [Bacteriophage 933W]
gi|4585436|gb|AAD25464.1|AF125520_59 percent identity 100
in 645 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail fiber protein"
/protein_id="BAB34651.1"
/db_xref="GI:13360688"
/translation="NSVYVSGTLKSPDGEAISGANTITALTIVSPDALSSTASAVTR
EGYIGTMDPEYAVSVTVKGTAVIGRVRIEGTSTVTLNMLRRSLVESIPGEL
LTDFIGQNNADDLATIRLNEDTATKNTQATSKESAAASAKSDSAKTISRAA
EAGQATDATEATRAVTAAGNAEESTRAGESEKAAGADAERQHAERKARLAQESA
GETLKRAEATVSAEAREAMENARGPRGPQGETGPKGDVPKGETGPPVQGPAGPK
GERGDVGAOGAVGAPGPRGEGPQGRGPQIPGLKGTGERGPKDGDGMPKPGSKG
DGGPAGPQGPKEGGEAGPQPMGARGERGTPRGEPPGPRGERGETGPPQGRG
EPGPASAAVATTAQGIVOLSSATSDDETKAATPRVAKAAMDVANEAKRAEE
AAGGVPQPKGDKGTGPAGPKDGRGDKGTGPVGTGRTGRTGRTGRTGRTGRTGRTG
KDRGERGTGLTGNAGPQGPQKGTGAAGPAGPQGPQGETGAAGPVGATGPPQPKGDP
GETQIRIFRLGPGNI IETNSHGWFPDPTDGLITGLTLPDKDATRVQGFQHLQVRFGD
GPMQDVKGLDEVGSDTGRTE"
149556..149825
/gene="ECs1229"
149556..149825
/gene="ECs1229"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585437|gb|AAD25465.1|AF125520_60
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34652.1"
/db_xref="GI:13360689"
/translation="MNILRKLQSLCGCKKHDDCENGQSILTAQLRLGPDADILESDENG
IPEQDRVITQWVILDADKKQIQCVVRPLQILRADGTWENIGMK"
149911..150153
/gene="ECs1230"
149911..150153
/gene="ECs1230"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585438|gb|AAD25466.1|AF125520_61
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34653.1"

gene
CDS

gene
CDS

gene
CDS

/db_xref="GI:13360690"
/translation="MLIRWSEGRVILVQEFFMFPENRRRIILDSKESWLIICDSQLGHL
MRSMYGRFRFLOLNLEKLGKGVHDVALPVKWEFTRQ"
complement(150040..150282)
/gene="ECs1231"
complement(150040..150282)
/gene="ECs1231"
/note="unknown, TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative outer, membrane protein"
/protein_id="BAB34654.1"
/db_xref="GI:13360691"
/translation="MLPFFHCIIIAAVATRSITPHTQDVSHSHHTVHFSPPPYTESVS
LSSCEPFFHWGDMYFPQFQIQLNKASLIHTAH"
150370..152073
/gene="ECs1232"
150370..152073
/gene="ECs1232"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585439|gb|AAD25467.1|AF125520_62
percent identity 100 in 567 aa, N-terminal part is similar
to hypothetical protein [Bacteriophage P-ElbD]
gi|7523538|gb|AAF63043.1|AF151675_5 percent 1"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34655.1"
/db_xref="GI:13360692"
/translation="MMPAIILANSIIINPLIFPEAVKGISMPIYIDITTRMRGMPRVVT
IRSPIAQDPHGRIYVTDGRFPKVTDAITATKGDGNHPTSSYRLGIDPAPTAPVCTVQO
SMLPESHVALEACHRFEGVITPEROISGVEKFTIKPKTIFHYRDOFWFANPDVVDV
GQVSDNDNDDETRFYTETFEVSDYGEEGPPGASLEVLTPTGTAQLTLAPVPLQN
ASTKRRIRYASGGGEADFLLVAELDASVLSYDKIPAKNLGSPSLATDMLPPEPNM
TGICLMANGIAAGFAGNEMFSEAYLFPAMPEVNRHTAEDIVAICPLGTSLYVATKG
EPIVSPQOSQNPASIVAYPWRGEVYACYTKPDGQDVFEVSPVNMDIRYLSITPFD
CAWVDLAKDMRVVTCGKMSVLGAGALPSTIRWHSKIFSLPERTSFCIRVKSAPER
VGTITWADDVPVVIHFAGTFKGSVVRLPATGQNVQWVMSGFGOVERITLTSNEMP
V"
152070..153338
/gene="ECs1233"
152070..153338
/gene="ECs1233"
/note="also weakly similar to tail tip fiber protein gp21
[phage N15] gi|7444604|pir||TI13107 percent identity 24 in
381 aa
identical to hypothetical protein [Bacteriophage 933W]
gi|4585440|gb|AAD25468.1|AF125520_63 percent identity 100
in 422 aa,
probable tail tip fiber protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail tip fiber protein"
/protein_id="BAB34656.1"
/db_xref="GI:13360693"
/translation="WTRKPRACKDLSTVVVENMEIGTGQRGDCGRHAFVTRRELVGLKL
ARRRTSGGASYALNPQIEIDSTLMTDFTPKPLNFATGGFSGVLLLEWDMPTNRGHSL
TEWRGTEDDLADAVLATTPGVQYDPPDQSGFYWIREFVNAAGYKGPWNAEKGTQ
AQTQIGVKAIIDQIRDEAAKSPVSELREIKNAQOAVKDAAIKTKTEVVTGREET
RTIGGTRISTLSDTSSELNEVDKRIKLDKEGGEAFKAWMSKKAGVDGITAGIGI
VAGKDEGRPVSVSAISASQLFVDPNNDNTAYPFAVSGGKYVVPKAMIYDAVETL
VSRKVVADVEKAGVSITSPVIRSAVIONGNFQVDSOGNLNIGGLSVTSQGLTIIRYS
NQNVGLVIRNDKTEVYDQNGRLAVIRGLR"
153353..153631
/gene="ECs1234"
153353..153631
/gene="ECs1234"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585441|gb|AAD25469.1|AF125520_64

/gene="Ecs1242"
/note="unknown, similar to hypothetical protein
[Bacteriophage 933W] gi|4585449|gb|AAD25477.1|AF125520_72
percent identity 99 in 2793 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34665.1"
/db_xref="GI:13360702"
/translation="WAYSEORPEAQLGNQRNSLNIOQGETDSYEAFFSPNRWKD
NSTPSLSGLVLTMGKGFPAQSVRGTEGMARGLGDMIQSPVTKGARILNEFRSMGLPG
VATVQDIFAGGSGADEVIDTLPDGNNAVTDIVGKLGKATKXAVSGAKATDEWLTKG
MSPGAVRALNTPMTEGYNDSVAVMKGVNLIGALVDMVAGGVARKVGDVTLRKMLTA
GLEKYIRAGMPQERATALAAEAVDKMPDLFQAGLITHTSVSAQCSAAADAVLN
ADYSELAQSPKFQOTFLSIDADPQHAQLTRDKMDLAKERVADEVRAQLATPELLVA
NMAAKLGLQLENLTVTRGTAKTVKSGIVRNATEOGAI NAAOGGYRYOENTALETA
GMGVSPWEGVADATIEGAAGAAGAPFCAGVAGYRGRRQAARETANRDAETVOQDDAA
POFESVDPVAQRESMOGNNRQLLEQYADADMAETEQAASAAHRRQAASOLLNELDEQ
TKRQAVMNLKAKPSLELLEYRLISQKEGRTETEQQOATREVI RPOQEVTPQAS
OPENAEDNGSIYPTVRFDPNEVRIEINGASPAERIEKVRPDNRYFTDEKSAMG
SDVFNAATGLKPSVVKKGNOYAVEMDNPAFSEDAVETINTLADGERIADADPME
OPAFWRDPFRGCTGDDTEVQARLAGRNAPTAEEIVRSOMAEAGDAGPTAQELTERPL
PAPGDIHPGCGYPLPGEVARTPDEQAAGRGRETTICEVKSQFQSQGAPAPENAAGR
QGETLEGDMVRRLGSPDAQNATAPVREGLPAPDIARNVMPQPESLPRTVRDLSPEL
AQAEEVROAGGNRDIPQETIAPSETTVDREATVRGGEVGRKKIEDFGEEIKGA
AKHRYAQLAETLKTLEDRYATQPLSKLPKPDYAKLANEGADADTLAMILYRSDI
PAKTHNTAGWGESIKKVHVSSEMLNGTVSAKRLAEWEGMPSRYADTWOLLRTLP
PSQMDRASARYVSVGQAAGKRYDPKOLYSLRNKDNKSGNLFESESDLELLAKAK
VWFAGEEKSQAQDEKTAPEDDKIRFDVYRNTRSGDIFIAYGNKNMVRGFGKFSAS
DARKYIDSHDELVRHVHEMREISREEQNATNRDTPERRKGNVSPQFSDAFGR
GVQFNGYEGPRQADLRAYDSLHDLAEVLNVPTKALSLNGLRLGAFAGARKGAAA
IFESGEVALNLTGKPGGALAHWFHSLDNFYGRYDVSDNGKITSGDGFMTAQRVRR
IKFGDYVDAEYVQOEYVDAFKYVIOAKNSMDPRRSALLDEVRSKPYWSTDVEMAA
RAFERYODKARMAGVENDYLVNIRKAPHEHNTDNTYAYPTNAELDGGIREADHLERT
LKTRETDKGVAYSRYSGVTRTPTEGNLISDVNKSAAKSGPVFOVAVARGVMSGIKDS
DLKRVYVSKQBAELAGELFDGVRGHAFYRDPDKREIVLVAADNIPDGRVTREKLRE
IITHAMEHYVTPAEYQITIIKVLKTRDSNVTIREAMRVKVDASYESPEVQAGEFLA
HMAEQKPNFVAAMERVALVKGVLRTTGLLKPTELNDLRLVRETIRTLGQRVREGYT
PREDDGAGSFQYSRSGKRPDPFKVPEGEGRYRDDLARMKSLRTDTLVNI GRTPPVL
RHLAGDPLVLSRDTVRKATNGVHKHVPMDVIERLPELMDHPDAIYRSATERNAVVM
LLDAVDKNGDPVVSVAHMKAVRSLEINKVASVYGTENGKKLKSMEMTGLTLYRREKL
SRDNLHRLGQLPKGEHSYRSADKILYPEDIRKGPYYISRTSLTPEETIASRFYQM
QDKFVLKAVQENIRKTKGQVDSNNAYMAEELFHGAENDLNVMKRYVQPLAKLLA
DYKIAQADIDELIYARHAPERNAHIAKINPKMPDGGSGMTNAAEAIMQVRVNSGKOA
OYDRLAGIIDDMLARRELIREAGLEENGVDQAWONAYRYVPLKGQDDVGVVSLPT
KGFTIGGRESQAMGRASQSPSQAIQDLSLIRHKNVEGNALFLKVQDNPK
DYQWFTDPRDPTMTIARCKQETGETIREVVERVPVMMADRTFTTKNGKTYII
KLHDPLRAMKSMGPTNSAEFVRLTGVNRFATVNTSYNPEFLVSFNIRDVQTAVM
NLKAEQGRSDGKLGLDNLISALAVKDSRSAMSAYASLRGKLTGKGQWQKRWKEF
VEDGGKTGFNMGDLGEOQKEMDRLVSLAKGGWKSGIGAWNSFLNLVEDANGAVENA
LRLSAYKHARDAGLSROOAAASLAKNMTVFNRRGEOGALMNSLYMFANASIOGTANLV
RTLHLNGSGPILLERLRWKNLAVPOKIALAAVAGYLLGSLNRSVAGEDDDGVNWDK
VPSHKVERNLVIMKSVYFGKAGEIWSIPLGYNTVFFLLHTAEGVAAAGDLTASRAAG
NVVGGQSLGRSTPEAYKAFASLWNLASGGSOYRPGAVDITPESLKFWDITISGPTF
GTPKPDQSLGRSTPEAYKAFASLWNLASGGSOYRPGAVDITPESLKFWDITISGPTF
RISKTQTAADVSLNGIDIPQEOVPFLKISGEVMPYADQOKMYDRMTETIAYHAELK
SLTGAERTAFIDENCKLSMGLMQDTRKRLDLRQDRDAIYADSTLSLAQQSAMVKS
VERDMKIADVRENREYNKKVGD"
167383. .167568
/gene="Ecs1243"
/db_xref="GI:167568"
/gene="Ecs1243"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] gi|4585450|gb|AAD25478.1|AF125520_73
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34666.1"
/db_xref="GI:13360703"
/translation="MFTPTVRRGTIPAIRADGTQIAHDDIDEEFFQVPLDGLFLISKYT

PFDIHALKDGVLQRTG"
complement(167519. .167584)
/note="I5609-01, with an internal deletion"
complement(167648. .167992)
/gene="Ecs1244"
complement(167648. .167992)
/note="Ecs1244"
[Bacteriophage 933W] gi|9632540|ref|NP_049534.1| percent
identity 100 in 114 aa, similar to hypothetical proteins
e.g. ygiw protein precursor [Escherichia coli]
gi|1739887|sp|P52083|YGIW_ECOLI percent identity"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34667.1"
/db_xref="GI:13360704"
/translation="MKKVLIAALISGVSFAGFAQQGFGQPEAERSTVAQAELKDDA
WVLEGSIVKVGDERYEFNRDSGTIVTIDDDSIWAGQNVSPDKVRIEGEIDKLSS
VEVDVKALKLLK"
complement(168112. .168324)
/gene="Ecs1245"
complement(168112. .168324)
/note="Ecs1245"
[Bacteriophage 933W] gi|4585453|gb|AAD25481.1|AF125520_76 percent identity 100
in 70 aa, also similar to Gelf [Escherichia coli]
gi|1786200|gb|AAC73129.1| percent"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Mokw protein"
/protein_id="BAB34668.1"
/db_xref="GI:13360705"
/translation="MLNTCLRSLVPRKQEKQAMQKQKAMLIIVITCLTVIVTALVT
RKDLCEVRITGQTEVAVFVYDYESEK"
complement(168558. .168953)
/gene="Ecs1246"
complement(168558. .168953)
/note="Ecs1246"
[Bacteriophage 933W] gi|4585454|gb|AAD25482.1|AF125520_77
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34669.1"
/db_xref="GI:13360706"
/translation="MENEGDNIITLVQPKRDEEKLNIITVTKRKNYTOQSCKHRAIEV
HEQDHVILCQCGVDPFOYVLRCAANDGEAVYREIRQLNRRHQDLRESVASLEREK
NTKARLFAARTALILYAENDUKNIEQKVNQ"
complement(168953. .169612)
/gene="Ecs1247"
complement(168953. .169612)
/note="Ecs1247"
[Bacteriophage V72-Sa] gi|5881668|dbj|BAAB4359.1| percent
identity 100 in 219 aa, identical to C-terminal part of
hypothetical protein [Bacteriophage 933W]
gi|4585455|gb|AAD25483.1|AF125520_78 percent identity"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34670.1"
/db_xref="GI:13360707"
/translation="MLELLDERERNQOYIKRRDOENEIEALTYGKLRVGLGAENNLII
DSECHVAELEELRQKALAESEKRIAELEALVSYTKLPHTQFQIYANLAEQMD
DGRTCAPHTDAQAEQNLACDGNRVOEYVKLERLQNALSGNSPVTDPGWSICSERMP
DTRKTVLVAVEFDKGDWRMKWATYIPGHPDANDGWIPGASWKPKSHWMLPEPPQEV

```
gene      N" complement(169633. .169851)
CDS       /gene="Ecs1248"
          complement(169633. .169851)
          /gene="Ecs1248"
          /note="unknown, identical to hypothetical protein
          [Bacteriophage VT2-Sa] gi|5881669|dbj|BAA84360.1| but
          different start, similar to hypothetical protein
          [Bacteriophage 933W] gi|7649907|dbj|BAA94185.1| percent
          identity 92 in 72 aa, also similar to hypothetical"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34671.1"
          /db_xref="GI:13360708"
          /translation="MSKIDYQALRKEAKATKGSYIVGHTSVNQHGNTGTVFCQKWK
          GEPGGVIAECHVNCIVETDAQAYATLNS"
          complement(169838. .170122)
          /gene="Ecs1249"
          complement(169838. .170122)
          /gene="Ecs1249"
          /note="unknown, identical to hypothetical protein
          [Bacteriophage VT2-Sa] gi|5881670|dbj|BAA84361.1| percent
          identity 100 in 94 aa, GTG start"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34672.1"
          /db_xref="GI:13360709"
          /translation="MNINTITIDTALNTGLALGVYIFMCSGRWLSLLFMKKWKNR
          RKOEQKAIDFAFFAEGDMEPGDPAIRSGVGVILYRSEKNEQD"
          complement(170119. .170340)
          /gene="Ecs1250"
          complement(170119. .170340)
          /gene="Ecs1250"
          /note="unknown, identical to hypothetical protein
          [Bacteriophage VT2-Sa] gi|5881671|dbj|BAA84362.1| percent
          identity 100 in 73 aa, similar to C4-type zinc finger
          proteins (Trar family) e.q. orf39 [Pseudomonas aeruginosa
          phage phi Cvx] gi|4063813|dbj|BAA36267."
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="C4-type zinc finger protein (Trar family)"
          /protein_id="BAB34673.1"
          /db_xref="GI:13360710"
          /translation="MADIIDNAAETELQRLNLSLQYKSDSNAPSATHCCECGDPIDE
          RRLAVRGCRACASCOQDIELINKRGVK"
          complement(170388. .171017)
          /gene="Ecs1251"
          complement(170388. .171017)
          /gene="Ecs1251"
          /note="also similar to putative phage anti-repressor
          proteins e.g. [Neisseria meningitidis]
          gi|7379969|emb|CAB84545.1| percent identity 49 in 112 aa
          identical to hypothetical protein [Bacteriophage VT2-Sa]
          gi|5881672|dbj|BAA84363.1| percent identity 100 in 209
          aa,
          probable anti-repressor protein,
          similar to hypothetical proteins H11422 [Haemophilus
          influenzae Rd] gi|1175795|sp|P44193|YE22_HAEIN percent
          identity 40 in 158 aa"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative anti-repressor protein"
          /protein_id="BAB34674.1"
          /db_xref="GI:13360711"
          /translation="MTSOLIPVFNGTIANETTLVNARDLHTFLGVGKRFASWITERI
          EBYGVENQDYTAISOKREIGYGRGKDYHLTLDTAKETAMVERNEKQGRIBRYIEC
          EKKLRNMQSAQEPQOQFTDEIILLCYNQLMWKRAQDLSKHLYPINKELNSSYNKL
```

```
gene      YDIAFETIYMVTKNRDVLLEAARDQSSFFVVRARPMKLSLRARQFEF"
CDS       complement(172166. .173560)
          /gene="Ecs1252"
          complement(172166. .173560)
          /gene="Ecs1252"
          /note="similar to YDGE_ECOLI gi|1787241 percent identity
          98 in 464 aa (Conserved in E.coli K-12)"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative transport protein"
          /protein_id="BAB34675.1"
          /db_xref="GI:13360712"
          /translation="MVRSQLSRADAPCLLTSLRQLMAMFGLPHWQLKSTSTESGVVA
          PDERLPFAQTAVMGVHAVAMFGATVPLIMGLDPNLSILMSGIGTLFFFITGGRV
          PSYLSSAAFGVGVIAATGNGQGINPNISALGIIACGLVYIVIGLVNWKIGTRWI
          ERLMPVTVACVYMAIGNLAPIAKSVSASGDFDSWNAVTVLCIGLVAVTRGMIOQ
          LLILVLGLIVACLLYGVMTNVLGKAVDFTLVSHAAMFGLPHFSTPAPNGOAMMLIAP
          VAVILVAENIGHLKAVAGMTGRNMDPYMGAFVGDGLATMLSGSVSGSVTVTYAENIG
          VMAVTKYSTLVFAAAVIAMLLGFSPFGALIHITIPAAVIGGASIVVFGLIIVAGAR
          IWQNRVDLSQNGNLIMVAVTVLVLGAGDEALTGLGFTLGGIGTATFGAILLALLSRR
          LVDVPPPEVVHQEP"
          complement(173515. .173973)
          /gene="Ecs1253"
          complement(173515. .173973)
          /gene="Ecs1253"
          /note="similar to B1007_ECOLI gi|1787242 percent identity
          98 in 152 aa (Conserved in E.coli K-12)"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34676.1"
          /db_xref="GI:13360713"
          /translation="MSCMGAANVITTDGAGRAGFTASAVCSVTDTPPTLLVCLNRS
          ASVWPFNERTLCVNTLSAGDEPLSNLFGKTPWEHFRFAARQWNGVTGCPGLEEAL
          VSPDCRISQVSVGTHDILFCAIEAIIHRHATPYGLVWFDRSYHALMRPAC"
          complement(174020. .174610)
          /gene="Ecs1254"
          complement(174020. .174610)
          /gene="Ecs1254"
          /note="similar to B1008_ECOLI gi|1787243 percent identity
          96 in 196 aa (Conserved in E.coli K-12)"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative enzyme"
          /protein_id="BAB34677.1"
          /db_xref="GI:13360714"
          /translation="MNEAVSPGALSTLTFTDARTINGWRETPTVSDETLRELYALMKWGP
          TSANCSPARIVFIRTAEKERLRPALSGNLQKTLTAPVTAI VAWDSEFYERLPQLFP
          HGDARSWFTSSPOLAETAFRNSMQAAYLIIACRGLGLDTPGMSGDRQHVDDAFA
          GSTLKSNNLINIGYDSSSKLFARLPRLSFEACGLL"
          complement(174620. .175420)
          /gene="Ecs1255"
          complement(174620. .175420)
          /gene="Ecs1255"
          /note="similar to B1009_ECOLI gi|1787244 percent identity
          97 in 266 aa (Conserved in E.coli K-12)"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative acetyltransferase"
          /protein_id="BAB34678.1"
          /db_xref="GI:13360715"
          /translation="MKLSLSPPYADAPVVVLLISGLGSGSYWLQPLAVLEQEVQVVC
          YDQRTGNNPDTLAEDY SITQMAAEHLQALVAAGIEHYAVVGHGALGVQMQLALDHP
          ASVTVLVCGNPNWLRINAHTRCFQVRERLYSGGAQVAEQPLFLYVADMMAARAPR
          LEAEDALALAHFOCKNNLLRNALKRADPSHHADRI RCPVQIICASDDLLVPSACSS
          ELHAALPDQCKVMRYGGHACNVTDPTPTFWALLNLGLSLHLHREAA"
          complement(175428. .175814)
          /gene="Ecs1256"
          complement(175428. .175814)
```



```

C-terminal part (37-276 aa in 276 aa) is similar to
B1017_ECOLI gi|1787253 percent identity 100 in 240 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cytochrome"
/protein_id="BAB34686.1"
/db_xref="GI:13360723"
/translation="MFVPELIMLREGLEAALIVSLTASYLKRTQGRGWIGMIGVLL
AAALCLGLGIFNETTGEFFQKEQELFEGIVAVIILTMVFWMRKVRNNKVQLE
QAVLSALQGNHGHWALVMVFAVAREGLSEVFFLLAQQDDVQWPIPLGLMGLAT
AVVLGELLWGGIRLNGAFFKWTSLFILFVAAGLAGAIRAPHEAGLWNHFOEIAPD
MSAVLSTHSLFGLTMEGIFGYQEPSVSEVAVWFIVLIPALVAFALPPRAGATASRA
"
188030..189157
/gene="ECs1264"
CDS
188030..189157
/gene="ECs1264"
/notes="similar to YCDO_ECOLI gi|1787254 percent identity
99 in 375 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34687.1"
/db_xref="GI:13360724"
/translation="MTINFRNALQLSVAALFSSAFMANAADVPQVKVTVTDKQCEPM
TITVNAKQTFIIONHSQKALEWEILKGVWVVEERENIAPGFSQKMTANLQPEYDMT
CELLNPKGLIVKGEATADAQSDALLSLGGATAYKAYVMAETQLVTDKAEFTDA
IKAGDIEKAKALYAPTRQHYRIETPAELFSDLDGSDIDAREDDYEQKAADPKFTGHR
LEKALFGNTTKMDQYADQLTVVDLQKRSIELAFPPSKVVGGAAGLIEEVAASKI
SSEEDRYSHTDLWDQANVEGSKIVDLLRPOLOKANPELLAKVDANFKKVDITLAKY
RTKDGFEYDKLTADRNALKGPITALAEDLAQLRGVLGD"
189163..190434
/gene="ECs1265"
CDS
189163..190434
/gene="ECs1265"
/notes="similar to YCDB_ECOLI gi|1787255 percent identity
98 in 423 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34688.1"
/db_xref="GI:13360725"
/translation="MQYEDKNGVNPSSRRRLKLGICALAGSCPVAHAQKTSAPGT
LSPARNKEQPYGEGHAGILTPQQAAMLVAFDVLASDKADLERFLRLTORFAFLT
QGGAAPETPNRPLPDLSDGILGGYIAPDNITITLSVGHSLFDERFLAPQMPKILQKM
TRFPNDSLDAALCHGDVLLQICANTQDTVIHALRDI IKHTPDLLSVRWKREGFISDHA
ARSGKETPINLLGPKDGTANPDSONDKLMQVWVWTDQOQEPANTIGGSYQAVRLIQ
FVPEWMDRTLPEKOOTIFGRKOTGAPLGMQHEHDVDPDYASDPKGLIALDSHRIAN
PRTAESSESLMRGYSISLGVNMSGQLDMGLLFVCYQHDLEKGLTVOKRLNGEAL
EYVRPIGGGYFFALPGVADANDYLGALLRV"
190778..191842
/gene="ECs1266"
CDS
190778..191842
/gene="ECs1266"
/notes="Similar to PHOH_ECOLI gi|1787257 percent identity
100 in 354 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ATP-binding pho regulon component PhoH"
/protein_id="BAB34689.1"
/db_xref="GI:13360726"
/translation="MVTSGTGHVLDNQRATTRGVFSSGSHLVTLHFQHPHFFSCVTDA
VNGARSFSAFYPRKAYGLOGSPDVRHAHRAAGACDEVKOLKVLMSGKQAVIKA
RREAKRVLRRDSRSHKOREESVTLVMQGVGEALGMARDSRDTSPILARNEAQLHYL
KAIESKQLIFATGAGGCKTWISAKAEALIHKDVDEIRIIVRPVLQADELDGLPGD
IAEFAPFIRPVDVLYVRUGDSFMQYCLRPEIKGVKEITAFAPYMRGRTFENAVVILDE
AONTYAQMKNMFLTRLGENTVTVIVNGDITQCDLPRGVCGLSDALERFEDEMVGIVR
FGKEDCVRSALCORTLHAYS"
complement(191892..192305)

```

```

/gene="ECs1267"
complement(191892..192305)
/gene="ECs1267"
/notes="similar to YCDP_ECOLI gi|1787258 percent identity
100 in 137 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34690.1"
/db_xref="GI:13360727"
/translation="MNNLIITRQSPVRLVVDVYVATILMTLTFALFELFAMDLLTGY
YWGSEARSRLQFYFLAVANAVLIWALYNKURFQKQHHAAAYQYTPQETASLAIP
DELYQLOKSHRMSVHTSQGIKMVYSEKALVRA"
complement(192307..193632)
/gene="ECs1268"
complement(192307..193632)
/notes="similar to YCDO_ECOLI gi|1787259 percent identity
99 in 441 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34691.1"
/db_xref="GI:13360728"
/translation="MINRIVSEFILCLVLCIPLCVAFVHSGELMRRFVFWPFMSIM
WIVGYVFWYRERWPGENAPOLKONPSIILPCFNEEKNVVEETTHAALAORY
ENETAVNDGSDTKTRAILDRMAAOIPLHRLVHLAONOGKATLKTGAANAASEVIL
CIDGDALDRDAAAYIVEPMYNPRVGTGNPRITRSTLVKGIQVGEYSIIIGLIK
RQRIYGNVFTVSGVIAAFRRSALAEVGSDDMIDTIDISWKLQNLQNTIYEPRRA
LCWILMPETLKLGLKORLWAOGAEVFLKNTRLMRKENFRMPLFFEYSLTIIWAF
TCLVGFIIYAVQVAFGLPNIETHIAATHTAGILLCTLCLLOFIVSLMIENRYEHNLT
SSLFIIWFPVIFWMLSLATTVSFTRVMLMPKQKQARWVSPDRGILRG"
complement(193625..195643)
/gene="ECs1269"
complement(193625..195643)
/notes="similar to YCDR_ECOLI gi|1787260 percent identity
99 in 672 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34692.1"
/db_xref="GI:13360729"
/translation="MLRNGNKYLLMLVSIIMLTACISQSRSTFIPQDRKSLIAEOPW
PHNGFVAISWHNVDEADQREFSVETSALRECFAMLRNGVQPVSIQRIAREHRGK
PLPEKAVLITFDGQSFYTRVPILOAQFQVAVWAPVGSVDTPADKQVKGDELVD
REYFATQQVREVARSLVELASHTWNSHYIOANATGSLLPVYVNRVYFTDHYET
AAEYRETRILDVAKMTEYLRTVEVNPVHFIWPGESQRIIMHIDLDYVVDVDMFTLES
GLANASQLDSIPRVLIANNPSLKEAQOITVQEKSPQRIIMHIDLDYVVDVDMFTLES
NIDVLIQRYKDMQISTVYLOAFADPDGDLVKEVFPNRLMLPKMADIFSRVWQLTR
SGVNIYANMPLVSLDLDPTLTRVKYLPTEKKAQTHPEQYHRLSPDDRVRQVGMLY
EDLAGHAFTDGLFPHDDALLSDYEDASAPITAYQAGFSGLSELRQNPEDFKQWAR
FKSRALTDTLELSARVKAIRGPHIKTARNIFALPIQPESEAWQADFLKSYDW
TAIMAMPYLEGVAEKSDQWLQNLQNKIPIQAKDKSILELAQANQWQNGHQALISS
QQALAHMSLLQNLGKNGYGYPDNLFNQHPEIDLIRPEFSTAWPKND"
complement(195652..198075)
/gene="ECs1270"
complement(195652..198075)
/notes="similar to YCDS_ECOLI gi|1787261 percent identity
100 in 807 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative outer membrane protein"
/protein_id="BAB34693.1"
/db_xref="GI:13360730"
/translation="WYSSSRKCPKTKWALKLLTAFLAASPAKSAVNNAYDALIIE
ARKQNTQALSWFALKSALSNQIADWLQIALWAGDQKVITVYRNRHQQLPARGYA
AVAVYRNLQWQNSLTLWQKALSLEPQNKDYQRGQIULTLADAGHYDTALVKLQLNS

```



```

KTVTAAAVIAKRYNTLLVHRTTELLKQWQERLAVFIQAGDSIGIIGGKHKPCGNID
IAYQSISSRQGEVEPLVRNGOIIIVDECHHIGAVSFSAILKETNARYLLGLTATPIRR
DGLHPITIMYCGAIRHTAVRPKESPHNLEVLIRSRFTSGHLPDSARDIQDIFREIALDH
DRTVAIEAAKFAFGQGRKVLVLTERTDHDLEIASVMNSLKPFIHLGRLSKKRRAM
LISGLNALPDPSPRILLSTGLIGEDFDHPLDPLTILAMPVSMKGTQLQQYAGRLHREH
TGKSDVRIIDFVDPAFYPVLLRMWDKRGYKAMGYRIIADGDESVI"
244276..244428
/gene="ECs1306"
CDS
244276..244428
/gene="ECs1306"
/notes="unknown,TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34729.1"
/db_xref="GI:13360766"
/translacion="MHNQFPFIQFPWFGLRNSQNFCHTIVVSGICTAFGWWYGNKCVF
LKITDI"
misc_feature
complement(244341..244724)
/notes="partial CDS, identical to L0015 [Escherichia coli]
transposase [Escherichia coli] gi|1552823|gb|AAB08675.1|
percent identity 100 in 81 aa"
repeat_region
complement(244481..247524)
/notes="IS30-1, truncated (5') and inserted by ISEc8-03"
repeat_region
complement(244726..247167)
/notes="ISEc8-03"
gene
complement(244763..246301)
/gene="ECs1307"
CDS
complement(244763..246301)
/gene="ECs1307"
/notes="unknown, identical to L0015 [Escherichia coli]
gi|3414883|gb|AAC31494.1| percent identity 100 in 512
aa, also similar to hypothetical proteins e.g. [Escherichia
coli] gi|3288156|emb|CAAL1509.1| percent identity 99 in
411 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34730.1"
/db_xref="GI:13360767"
/translacion="MNDISSDDIFLLKORLAEOALHAEKLSNPEREIDLHQAOL
DKLRNMFGSRSEKVSRIQAEDNLRLQKESDITLGRVIDPAVQRLQTRTRKPF
PESLDPDEKRLPAPCCPGGSLYGEDTAELQELMSAFRIVRTREKHACTQC
DAIVQAPASRPRIERGIAGPLLARVLTSKYAETHPLRYQSEIYGRQVELRSLSG
WSDACCLSPLEALHGYYMTDGKLHADDTPVQVLLPGNKKTKTGLWAYVRDRNA
GSALAPVAFYSPDRKGIHPQTHLACFSGVLOADAVAGNELYRNGGITEAACWAHA
RKIHVDVRIPSALTTEALEQIGOLYAIJEAIDIRGMPAEOFLAEROKTKPLKLSLES
WLREKMKTLSRHSELAKAFAYALNQWPAITYYANGWVEIDNIAENALRAVSLGRNA
FLFFGSDHGGERGALLYSLTGCKLNDVPDESILRHVLGVIAOWPVRVSELLPWRIA
LPAE"
gene
complement(246351..246698)
/gene="ECs1308"
CDS
complement(246351..246698)
/gene="ECs1308"
/notes="unknown, identical to L0014 [Escherichia coli]
gi|3288157|emb|CAAL1510.1| percent identity 100 in 115
aa, similar to hypothetical proteins e.g. orf50
[Escherichia coli] gi|6009426|dbj|BAA84885.1| percent
identity 76 in 107 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34731.1"
/db_xref="GI:13360768"
/translacion="MIFPLSGTKILWVAGITDMRNGFNGLAAKVQTLAKDDPMHGHVF
IFRGRGSQVKLWSTGDGGLCLTLKRLERGFAMPARGDKVELTQAQAMLLLEGIDW
RQPKRLTSLTML"
gene
complement(246695..247096)
/gene="ECs1309"
CDS
complement(246695..247096)

```

```

/gene="ECs1309"
/notes="unknown, similar to hypothetical proteins e.g. L0013
[Escherichia coli] gi|3414881|gb|AAC31492.1| percent
identity 98 in 133 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34732.1"
/db_xref="GI:13360769"
/translacion="MGTKVSMDQKNVTPGRKKGCPNPPEFKQOLVAASCEPGISISK
LALENGINANLLFKWQOOREGKLLLPSSSQPLLPTVLDAAAEQPSLAEDPEPLSI
SCEVTFRHGTGLRFNGVSEKLLTLIIQLKR"
complement(247151..247462)
/gene="ECs1310"
CDS
complement(247151..247462)
/gene="ECs1310"
/notes="IS30 transposase (interrupted), similar to
N-terminal part of IS30 transposase
gi|2851554|sp|P37246|TRA8_ECOLI percent identity 99 in 101
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="IS30 transposase"
/protein_id="BAB34733.1"
/db_xref="GI:13360770"
/translacion="MRRTEAEKASVFELWKNGTFSEIANILGSKPGTIFTMLRDT
GGIKPHERKRAVAHLTLSEEREIRAGLSAKMSIRAITALNRSPTISREVRNRKRT
A"
247523..248463
/notes="IS600-1, truncated (5')"
gene
247602..248420
/gene="ECs1311"
CDS
247602..248420
/gene="ECs1311"
/notes="probable transposase, similar to transposases e.g.
Hpl [Escherichia coli] gi|3661482|gb|AAC61713.1| percent
identity 98% in 272 aa, InsB [Shigella dysenteriae]
gi|5532467|gb|AAD44751.1|AF141323.22 percent identity 98
in 272 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase"
/protein_id="BAB34734.1"
/db_xref="GI:13360771"
/translacion="MCQVFGVSRSGYVNWQHEPSDRKQSDERLKLKIKVAHIRTRET
YGRRTQTELAENGIIVGRDLARLKRKLRCQKRRFRATTNSNHNLPVAPNLLNQ
TFAPTAPQWVADLTIVATQEGWLYLAGIKDVTTCIIVGYAMGERMTKELTKALFM
ALRSQRPAGLIHHSDRGSQYCAIDYRIOEQFLKTSMRKGCYDNAPMESFVGTL
KNESLSHYRFSRDEALISVIREIETIYFNQRQRHSLGNISPAAFREKYHQMAA"
248326..249072
/gene="ECs1312"
CDS
248326..249072
/gene="ECs1312"
/notes="probable complement resistance protein
precursor, similar to lipoprotein trar precursors e.g.
gi|418135|sp|P32885|Tru1_ECOLI percent identity 83 in 227
aa, TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative complement resistance protein
precursor"
/protein_id="BAB34735.1"
/db_xref="GI:13360772"
/translacion="MRFSTIVSVTLVWGSIPROPSGKNITIRLLKKRTNGSVRYCOY
TSTIWLPEASRETVFLOIKNTSDKMSGLOGKIADAVKARGYQITMSPDKATYWIQAN
VLKADMLRSESQWLSRGVEGAAGAALGGGITYAVNTSAGTTLGLGLATYIGMAA
DAMVEDINTMITDVOIARTRTSVRDNVAAALRQGTSGSKIQTSTETGNQHKYQTRV
VSSANOVNLKFEKAPHELDOLAKSIANIL"
complement(249120..249371)
gene

```

CDS
 /gene="ECs1313"
 complement(249120. .249371)
 /gene="ECs1313"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34736.1"
 /db_xref="GI:13360773"
 /translation="MESFVQDSPFYSGRDLYWLRPKVELTLEEKLYCSCIRNRHKY
 SVGRQARNTLKNLLYPSLDSVPAWYGVYTKIISELSE"
 complement(249395. .249685)
 /gene="ECs1314"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34737.1"
 /db_xref="GI:13360774"
 /translation="MNLASMDNADGDFELMSGNQDNICCRDNRHYPKHKLPLS
 ELFDVNYGUNLEKLEKSDSGINFVRSYTKNGVSARVKLKDGTSPCOOLC"
 complement(249761. .249961)
 /gene="ECs1315"
 complement(249761. .249961)
 /gene="ECs1315"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34738.1"
 /db_xref="GI:13360775"
 /translation="MIDADLSRYAQLWTQIGIFSARYSYQIDDLFAQMLFVLSQPGFC
 GHHCSPMGNRDRRLAVMRFCR"
 complement(250371. .250730)
 /gene="ECs1316"
 complement(250371. .250730)
 /gene="ECs1316"
 /note="probable diacylglycerol kinase, similar to
 diacylglycerol kinases e.g. gi|125321|sp|P00556|KDGL_ECOLI
 percent identity 76 in 119 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative diacylglycerol kinase"
 /protein_id="BAB34739.1"
 /db_xref="GI:13360776"
 /translation="MANTTGVTRIKAAGYSWKGLCAAWRYEAAFRORCILVLSALI
 ISFILDVSVTRVLLIGSAMVIVIEIINSAEAVVDRIGSEHHEUSGRAKDMGSAAV
 LITILVAIIWGVLILLS"
 complement(250823. .252109)
 /gene="ECs1317"
 complement(250823. .252109)
 /gene="ECs1317"
 /note="unknown, similar to hypothetical proteins e.g. yjdB
 in bss-adiy intergenic region [Escherichia coli]
 gi|731986|sp|P30845|YJDB_ECOLI percent identity 45 in 428
 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative outer membrane protein"
 /protein_id="BAB34740.1"
 /db_xref="GI:13360777"
 /translation="WTSYFNMSVWLWFLTGILPGCLLLVNLIRYPETWIKGITVRLI
 SWFASLLIIFAIAPFYKDYASVGRNNSLNKEIPTNYIYSGFYKVRFFYSPGFR
 QTGDASRTINEKQRPVIMFLVVGETARSONALNGISRGTDFTKKYNELISFHNQV
 SCGTSTAISVPCFM SDMKRKEFNKRVNSENVLDILYRTGVNLLIENDGGCKGVCK
 RIPTINIEPSNDTLCNKNKSCYDEVMLKNIDEYINNSEDKLIIVFHLMSGHPTYYL

RYPESHKYFKPTCDRSDRIENCTHEQLINTYDNTIRYTDYIISKLDKLEKDEYDTV
 LLYVSDHGESLGENGLYHGTPYNVAPAQTHVPLITWSPGFVSSKKIDLNCLESHA
 LNRVTSHDNIFSLGLGNWNTSVYNKEDDIISCR"
 complement(252369. .252512)
 /gene="ECs1318"
 complement(252369. .252512)
 /gene="ECs1318"
 /note="unknown, TTG start"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34741.1"
 /db_xref="GI:13360778"
 /translation="MNVAKTTSIINIKRKLILIFLIILPLIRAYNCPYLKVIVFY
 KTW"
 complement(252667. .252942)
 /gene="ECs1319"
 complement(252667. .252942)
 /gene="ECs1319"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34742.1"
 /db_xref="GI:13360779"
 /translation="WTEHWSGHADDNFPFSWGSDSIFMHACLSQSYGESICTVHTFVA
 DVAPESLTGLQTVYLYTAGLLCDIVFHEIGGPEPMNGIIFPDIIIP"
 complement(252994. .253197)
 /gene="ECs1320"
 complement(252994. .253197)
 /gene="ECs1320"
 /note="unknown, GTG start"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34743.1"
 /db_xref="GI:13360780"
 /translation="MSFRLNNTNVCFRHEAFRLPGLYIARKAMHIYGIITQRQRDPT
 MGDYAIWLRMISRCYGDNLLIQ"
 complement(253278. .254021)
 /gene="ECs1321"
 complement(253278. .254021)
 /gene="ECs1321"
 /note="urease accessory protein Ured, similar to Ured
 urease-associated proteins e.g. [Klebsiella aerogenes]
 gi|731078|sp|Q09063|URED_KLEAE percent identity 71 in 242
 aa, TTG start"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="urease-associated protein Ured"
 /protein_id="BAB34744.1"
 /db_xref="GI:13360781"
 /translation="MLSTALPONKKGQAMLDLRFORLHGKTTLTTRHHVGLITVORP
 FPEETCHLLHPGGIVGDELINAIIDSDCHTLITMPGASKFYSSGQAHLQ
 QNLTCNNATLEWLPQDSIFFPGAHAHTLVHFLSSSTLLAWDLCLGRPVIGETFS
 HGTLANRLTWGDSPLLIERLHVANGELTVCARKPWVGTMFLYLGNETQLEDIREKL
 TPLENYAGATLTDGLLTVRFLSDNLRQC"
 complement(254112. .254414)
 /gene="ECs1322"
 complement(254112. .254414)
 /gene="ECs1322"
 /note="urease gamma subunit, similar to urease gamma
 subunits e.g. [Klebsiella pneumoniae]
 gi|137084|sp|P18316|URE3_KLEAE percent identity 96 in 100
 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="urease gamma subunit"

```

/protein_id="BAB34745.1"
/db_xref="GI:13360782"
/translation="MELTPREKDKLLFTALLAERLARGLKLNYPESVALISAFIM
EGARDGKSVAALEMEGRHVLREQVMGIPMDIIOVEATFPDGSGLVTVHNPDI"
254423..254743
/gene="ECs1323"
CDS
/gene="ECs1323"
/notes="urease beta subunit,similar to urease beta subunits
e.g. [Klebsiella pneumoniae]
gi|137077|sp|P18315|URE2_KLEAE percent identity 82 in 106
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="urease beta subunit"
/protein_id="BAB34746.1"
/db_xref="GI:13360783"
/translation="MIPGEYKVRPGYIELNIGRATCSIIIVENHGDRPIQGVSHVFAE
VNPALKFDQKARGYRLNIAAGTAVRPEPCQKREVELVALSGARIVHGFGRDINGELE
ANDE"
254733..256439
/gene="ECs1324"
CDS
254733..256439
/gene="ECs1324"
/notes="urease alpha subunit,similar to urease alpha
subunits e.g. [Klebsiella pneumoniae]
gi|137070|sp|P18314|URE1_KLEAE percent identity 90 in 567
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="urease alpha subunit"
/protein_id="BAB34747.1"
/db_xref="GI:13360784"
/translation="MMSNISRQAYDMFGPTGDKIRLADTELWIEVEDLTYGEEV
KFGGQVIRDMGQGMQLSAGCADVLVTNALIIDYWGIVKADIGVKGDRIFAIGKAGN
PDQPNVTPIGVSTETIAAEGRIIVTAGVDTHIWLCPQAAEALTSGITTMIGGT
PTAGSNATVCTPGPWIXOMLOADSLPVNIGLLGKNCNPDALREQVAAGVIGLK
IHEWGCATVATCAEDMTADMDVQVALHSDTLNIESGFVEDTLTAIGGRTIHTFTEG
AGGHPADIITACAHENILPSSNTPLPYTVNTIDHLDMLVCHHLDPDIAEDVAF
ESRIGETIAEDVLHDLGASLTSSDSQAMGVEGVWLRTQVAHRMKVQGPLPEE
SGDNVNRVKRYIAKYIINPALTGHIAHEVGSIEVGKLADLWLSAPFVGPATIVK
GMIAAMPMDINGSIPTPPVHYRPMFAALGSAHRRCRVTLFSSAAANGVAEQNLN
HSTTAVVKGCRTVQKADMRHNSLLPDIITVDSQTYEVRINGELTISEPAILPMAQRYF
LP"
256449..256913
/gene="ECs1325"
CDS
256449..256913
/gene="ECs1325"
/notes="urease accessory protein,similar to UreE urease
accessory proteins e.g. [Klebsiella aerogenes]
gi|137095|sp|P18317|UREE_KLEAE percent identity 80 in 154
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="urease accessory protein UreE"
/protein_id="BAB34748.1"
/db_xref="GI:13360785"
/translation="MLYLTRRVTPAQTTASVTLPVDMRVKSRKIVTLNDGROAGILL
PGLLLRDGDIILSNENGDEFIKVIAADEAVSVRCADPMLAKACHWILGNRVPLOIM
PGELRYHHDHVLDDMLRQGLDVDFAHLPFEPEAGAYASKSHAHNDQESH"
256914..257588
/gene="ECs1326"
CDS
256914..257588
/gene="ECs1326"
/notes="urease accessory protein,similar to UreF urease
accessory proteins UreF e.g. [Klebsiella aerogenes]
gi|137097|sp|P18318|UREF_KLEAE percent identity 79 in 224
aa"
/codon_start=1
/transl_table=11

```

```

/evidence=not_experimental
/product="urease accessory protein UreF"
/protein_id="BAB34749.1"
/db_xref="GI:13360786"
/translation="MPTPEKRLMQLASNSLPVGYSSWQGLEWAVEAGHVDESAAF
EHMQLOMEQSFAYDLPLLARLYRAGEAGDPDSAGRWATYLLACRETRERDEERNR
GAGTFLVDMQPCPAENWRKLCCQSQLTQAMWLGVRWQISVDLALSLSGYSWIESAV
MAGVRLVPIYGOLAAQQLIMRLICARYAANMDSALATPDPAIGSATPLASIASARHETQY
SRLFPS"
257600..258217
/gene="ECs1327"
CDS
257600..258217
/gene="ECs1327"
/notes="urease accessory protein,similar to UreG urease
accessory proteins e.g. [Klebsiella aerogenes]
gi|137099|sp|P18319|UREG_KLEAE percent identity 90 in 205
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="urease accessory protein Ureg"
/protein_id="BAB34750.1"
/db_xref="GI:13360787"
/translation="MNIKQPLRVGVGPGVSGKLTALLEALCKSMRDTWOLAVVTNDI
YTRDQRILTEAGALEARNRIVGTCPTAIRFEDSMNLAAVEALSERFNGLELIF
VESGONLSATTFPELADLTYYIDVAEGEKIPKGGGITKSDFLVINKTDLAPYVG
ASLEVMRDTLRMRGERPMGFTNLKSGEGLQNIATFEEQCMILGK"
complement(258265..258546)
/notes="partial CDS, similar to C-terminal parts of
autosecreted serine protease e.g. Pet[enteroaggregative
Escherichia coli] gi|3095184|gb|AAC26634.1| percent
identity 92 in 79 aa"
complement(258473..258814)
/gene="ECs1328"
complement(258473..258814)
/notes="unknown,similar to hypothetical proteins e.g. TnpJ
[Shigella flexneri] gi|5532468|gb|AAD44752.1|AF141323_23
percent identity 100 in 87 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34751.1"
/db_xref="GI:13360788"
/translation="MSRKTQYKSEKFAEAVRTVLENQLSISEGASRLSIPGTLGQW
VTAARKGLTGPGRSRTVAELESITLQRLKALNEARLERDILKKAADVSICQNTPFNVL
FHCPRRYRLSV"
258505..258551
/notes="IS30-2, truncated (5')"
complement(258552..258878)
/notes="IS600-2, truncated (3')"
258853..259008
/gene="ECs1329"
258853..259008
/gene="ECs1329"
/notes="unknown,GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34752.1"
/db_xref="GI:13360789"
/translation="MSVKLRLPQVSLIWFVTSRFRFLIAPVPMAPHERASAEITSAII
ITFLVFI"
259429..259692
/gene="ECs1330"
CDS
259429..259692
/gene="ECs1330"
/notes="GTG start
ribosomal protein L31-like protein,
similar to hypothetical proteins e.g. ribosomal protein
L31 homolog ykgw in intf-eaeH intergenic region"

```

[Escherichia coli K-12] gi|3025204|sp|P71302|YKGM_ECOLI
percent identity 93 in 86 aa
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ribosomal protein L31-like protein"
/protein_id="BAB34753.1"
/db_xref="GI:13360790"
/translation="MKPNIHPEYRTVVFHDTSIDEXFRIGSTIKTRVIELDGVTVPY
VTIDVSSKHPFYTGKRLTVASEGNVARTQRFGRFVDAKKG"
complement(260066. .260347)
/gene="Ecs1331"
complement(260066. .260347)
/gene="Ecs1331"
/note="unknown,GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34754.1"
/db_xref="GI:13360791"
/translation="MKFLTESSAPFYGAFFGAPTLLPHFRDANKKAHSEYTDGCFFF
VICVQCDRPNLWQPSGDCPCRCFLRTLQERNFRPEAYPEFR"
complement(260619. .260960)
/gene="Ecs1332"
complement(260619. .260960)
/gene="Ecs1332"
/note="probable colicin immunity protein, similar to
colicin immunity proteins e.g. colicin E1 immunity protein
gi|1243951|sp|P02985|IMM1_ECOLI percent identity 25 in 107
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative colicin immunity protein"
/protein_id="BAB34755.1"
/db_xref="GI:13360792"
/translation="MNRKFYFRNLWGMCPPIISILHLIYDPTPEASIKVDLFLISG
LALPCSIYITKTTILKYSITSEFWKPLFWSYGIYHFLVVICIVLAIPSAIIVLFF
HFYONITRNWK"
complement(261006. .261677)
/gene="Ecs1333"
complement(261006. .261677)
/gene="Ecs1333"
/note="unknown, weakly similar to hypothetical protein
MAL4P2_26 [Plasmodium falciparum]
gi|6562728|emb|CAB62867.1| percent identity 29 in 106 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative membrane protein"
/protein_id="BAB34756.1"
/db_xref="GI:13360793"
/translation="MEFSVIFFWITGLTILFLKSIITDLINKLNKTDIALYILAPL
SAGYQYIYNDFSELGWSGLFVILNIVNLFLLFLSKMLMSHRIKMQ
KKLKQTSVPDFIYSEIKHKKYPIYVLHPSNNIIEIGNIINTDLVIGEKIHFRTL
NRSLYFNLTHPLIATIKDDVFCTLHDYNNHNETKSTIDNIYSKIKDNHNTWLLNN
ENVQE"
complement(261880. .262035)
/gene="Ecs1334"
complement(261880. .262035)
/gene="Ecs1334"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34757.1"
/db_xref="GI:13360794"
/translation="MYHHKGILFQATRILOSPGVSIRLLKHWLAVTRRRPSQTAGTL
ACRKSDS"
262752. .263513
/gene="Ecs1335"

CDS
262752. .263513
/gene="Ecs1335"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34758.1"
/db_xref="GI:13360795"
/translation="MHTNDRSLTMVDLFCGAGATARLVSEAGSEELRYCSELILTHF
SGEPLAAHTIDHRLTMVDLFCCTGATARLVSEAGSEELSRVSIASGRVLTFRVLMAQDE
VQPEWLRADSGDSLSARVVAALAHFLSSGVQLPTAGYGLHEHLWAGLTHFEKVP
SGNSSSSSSGHSRRVTSVQACRPLRFVAVSRRLIICAAFPFAISEPANNHAF
RPMTTGRIARSKQLLSRGTAPELRKQLSSRRFRF"
complement(263210. .263809)
/gene="Ecs1336"
complement(263210. .263809)
/gene="Ecs1336"
/note="unknown, similar to C-terminal parts of hypothetical
proteins e.g. [Pseudomonas putida]
gi|2995633|gb|AAC98738.1| percent identity 40 in 200 aa,
L0015 [Escherichia coli] gi|3414883|gb|AAC31494.1| percent
identity 39 in 200 aaGTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34759.1"
/db_xref="GI:13360796"
/translation="MYKSLADNDPGTTSACMWSHARRGFANLYKASREPRAMALRKI
AGLYRIEKFIREFRVEKIRQWRYSRIVNDFLFWPEQPCPPDGLNKAINYL
NRRDELSCLFDGAVPLDNNICERAIRPVVMGRKAMLFAGSLMAGNRAAQIIMSLLETA
KRNGLPHAWLTDVLTRELPEWEDRLEELLPLEGTFSG"
263929. .266461
/note="I5682.1"
264015. .264440
/gene="Ecs1337"
264015. .264440
/gene="Ecs1337"
/note="unknown, similar to hypothetical proteins e.g. orf29
[Escherichia coli] gi|6009405|dbj|BAA84864.1| percent
identity 37 in 136 aa, L0013 [Escherichia coli]
gi|3414881|gb|AAC31492.1| percent identity 38 in 124 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34760.1"
/db_xref="GI:13360797"
/translation="MEQKILSSEPRRSFNEFKLQMYKLASQPGASVARIAREHDIND
NLLFKWLRLWQERRISRRLPVTSSGAGVELLPVEITPDQKEPMAALTPLLSTPSQ
STVSASSCKVEFRHGNMTLENPSPELLTVLIRELTGRGR"
264437. .264787
/gene="Ecs1338"
264437. .264787
/gene="Ecs1338"
/note="unknown, similar to hypothetical proteins e.g. orf50
[Escherichia coli] gi|6009426|dbj|BAA84885.1| percent
identity 71 in 106 aa, L0014 [Escherichia coli]
gi|3288157|emb|CAA11510.1| percent identity 64 in 116 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34761.1"
/db_xref="GI:13360798"
/translation="MISLPSGTRILWLVAGVTDMRKSFNGLGEQVQHVLDENPFSGHLF
IFRRRGDTIKILWADADGLCTKRLEEQFIWPAVRDGVKSVITRSQSLAMLKLDKLDW
ROPKTSRLNALTML"
264818. .266431
/gene="Ecs1339"
264818. .266431
/gene="Ecs1339"

```
/note="unknown,similar to hypothetical proteins e.g.  
L0015 [Escherichia coli] gi|3414883|gb|AAC31494.1| percent  
identity 61 in 503 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34762.1"  
/db_xref="GI:13360799"  
/translation="MSRKYLIRITELERLISEQAEALRQDQSLVETEAEFLRSAL  
ARAEEKTEERELEHLRAQIEKRLMLFGT RSEKLOREVEQAEALQKREQESDRYS  
GREDDPQQLRQSRHRRPLPAHLPEIRHLEPEESCCPECGSELVLCGEAEQLE  
LVSAALKVITRVYKAKTCDICVEAPAPSPRIERGICGGLLRLVLTCKYCEHLPL  
YRSEIEFAROGIELSRALLSNWVDACQMLTLLNDTLIRYVMNTRKVHTDDTPVKVLA  
PGRKKATGRIWTVYRDRNRAGSEPPAVFYAPSQGRGHPVOHLRPFPGIQQADAF  
SGDYRLSAREGALTEVACMAHARKV IHDVYISSKSATAEALKRISILYAEDEI  
RGYPESRLAVRQORSKVLTSLEHMVVERNGTISKSRIGCEACSYVLNQWDALCYYS  
DDGLAEDNNAARALRAVCLGKKNFMFFGSDHGGERGALLYGLIGTCRLNGIDPEAY  
LRHLSVLPWPSNRVDELLPWNVLTNK"  
/complement(266435..267163)  
/gene="ECs1340"  
CDS  
complement(266435..267163)  
/note="ECs1340"  
/note="unknown,similar to hypothetical proteins e.g. 57.8  
kD protein [Pseudomonas putida]  
gi|2496740|sp|P55630|Y4Q1_RH1SN percent identity 37 in 232  
aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34763.1"  
/db_xref="GI:13360800"  
/translation="MKLARQREFKCKESLAGMQRSLFEEDVDADIAEISAHLDKLLP  
QTQDEKTTTPRVKPLPSPLPRAEKVIPPAAERCPDCDAPLHFIRDEVSKLEYIPA  
QVVNVYIRPOYSCPCCKEKFVSGKMPAHILPKSAVEPSVIAQVVISKYTHPLPYRQ  
HIFSRMGVELPVSTMADMGVAGALAPLAKLLRHELLTRDVIHADETSRLIDTRKG  
GKSCGHLNRYVSGERSVNGAY"  
/complement(267374..267736)  
/gene="ECs1341"  
CDS  
complement(267374..267736)  
/note="ECs1341"  
/note="unknown,similar to hypothetical proteins e.g. 20.3k  
protein [Agrobacterium tumefaciens Isl131]  
gi|95090|pir||JC1151 percent identity 41 in 101 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34764.1"  
/db_xref="GI:13360801"  
/translation="MFSGLFAMLPDNLVFLVVKPVDMMRGIDLTQYVQNELNAAWHD  
GAAFVFTNKVRSRIKVLRWDDKHGWLCRLRLHRGSRFWPRKGDATWHLTQDEPHLVF  
GVDWQOVKGHDIAKWVYQ"  
/complement(267702..268031)  
/gene="ECs1342"  
CDS  
complement(267702..268031)  
/note="ECs1342"  
/note="unknown"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34765.1"  
/db_xref="GI:13360802"  
/translation="MSKPRWTLDDKKHVAANRASGLTREQYCELYDIPFKSIRWMPQ  
DVAKAEKRARAPEIIPVSVSGSGMTDGRPLSDPEVTLPLPGGIMCCQPSQLTDVFR  
ALRHADA"  
/complement(268292..268759)  
/gene="ECs1343"  
CDS  
complement(268292..268759)  
/note="ECs1343"  
/note="immunity to R478 phage/colicin/tellurite resistance
```

```
cluster,similar to TerW [Plasmid R478]  
gi|1354147|gb|AAC44736.1| percent identity 99 in 155 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="TerW protein"  
/protein_id="BAB34766.1"  
/db_xref="GI:13360803"  
/translation="MOLNTRQARIFKLANLIGTGKPVSAADIITSLSECSEPTLRLAKL  
ELRYSAGETIKYSKAGSYHLVNPQOLDKKTLRMNEALQAQNAELKTGEGTGKVVLDK  
DKTAVLSLRMRTLRIKIDRLAALSDSTRSEAVEKLAHSHVDELIKEYSAKKS"  
/complement(268777..269985)  
/gene="ECs1344"  
CDS  
complement(268777..269985)  
/note="ECs1344"  
/note="unknown"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34767.1"  
/db_xref="GI:13360804"  
/translation="MGPRFRKSNILPGVRLNLSNGAPSLSVGPRGASVFGSRGTVA  
NLGIPGTGLSYRTLDRAARGSENTATDPGLRQALEQKAAELMSAVTAIRNIHELIT  
PDKTIGISWAELEAVYLHNRTSPFQVPAPVRPEKPDYLVLPKPAESEGISFLGKWFPE  
SESAKAEARHAENLRWQELIDVERENTLROHRYQQOQTAWAEQYANWKFEAEHEKER  
LATAQADARQOFRDAAFFESYLAGVLAETEMPRETIVAEVPELSAVLLDVLDAEI  
EDFPDKIYGVNARGTETEKAMQKAVRENYAHHVHGCLFRLVGLVLTLPFDNVIVS  
GFTQVSKRTGYLDEYLLSCKCTRSQSSVNFAGIKHIDPVEALGDPQVIRKMSSTF  
IFQPIEPLTL"  
/complement(269996..270952)  
/gene="ECs1345"  
CDS  
complement(269996..270952)  
/note="ECs1345"  
/note="unknown,similar to hypothetical proteins e.g.  
[Deinococcus radiodurans] gi|747167|pir||B75302 percent  
identity 42 in 305 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34768.1"  
/db_xref="GI:13360805"  
/translation="MKNRLSPNMGATLYMPATREDIADAVLHGKIPGLRSLVICLED  
AVSEADIPVALKNEHLHLLSLSMSGLKNDWPLVFI RPRHAEIMGRLAKHAYDLSAV  
DGEVLPKFTLSLAEWMDIMVGTIHLCMPTLETEDVVDVQMRRLATRLHEEPCHDR I  
IALRIGSNDLMNVVSLRRPDLTLYDSPMGYVILKLVSVFGPRDEALTAPVCEHIDDH  
AVNARELALDMAHGLVKGTAHFEGQIEVQNALMVTQGEHSDALRILNLTQAVFKSOG  
AMCEPATRRWAAAGIILDRARFYQLQNEQSDAGIRLLTUTVTOHH"  
/complement(270952..271686)  
/gene="ECs1346"  
CDS  
complement(270952..271686)  
/note="ECs1346"  
/note="unknown,similar to hypothetical proteins e.g.  
[Streptomyces coelicolor A3(2)] gi|742048|pir||A75302  
percent identity 34 in 260 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34769.1"  
/db_xref="GI:13360806"  
/translation="MGKTSWHYGISITIRDRGIDGAALDVIIERHGTSGTIVFDVGTGK  
GAITGELVALKDRPCYEQPRLVVLADPCGSWLAASDDDLIPGIMGAPVSGILIS  
RSVWSSEGHLGCMVCHEHLSFEFCSRMLADTVAFHFKLTPSSLAPLSWNTESARYLWQ  
TSRDVIAFLADEKVDVSNRIKPGIAENRAVLRRVPDHFVFRKSIDDDPVALLVGLAR  
EKGIIVTEMGGTIGQYRAVYTIKKVL"  
/complement(271688..272068)  
/gene="ECs1347"  
CDS  
complement(271688..272068)  
/note="ECs1347"  
/note="unknown,similar to hypothetical proteins e.g.  
[Streptomyces coelicolor A3(2)] gi|8246803|emb|CAB92838.1|
```

percent identity 45 in 97 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34770.1"
/db_xref="GI:13360807"
/translation="MPFSEFLEKIKMQPHETFTGSYPQGDVEFLKPPVVIEMTPVEK
KEELIQSGKKHYSMDLSPQAPTOHWLDFLFRALDRGAERLAKEVQTALALAEAFGD
EPTVLASLVRAGVPLGRYAAPGA"
complement(272034. .272807)
/gene="ECs1348"
CDS
complement(272034. .272807)
/gene="ECs1348"
[note="unknown, similar to hypothetical proteins e.g.
[Serratia marcescens] gi|1695868|gb|AAB37122.1| percent
identity 100 in 167 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34771.1"
/db_xref="GI:13360808"
/translation="MNKPVVLSDLDTLFQTRKKMVDELALFRTGAVDRTLNPRSF
MTEQSMVLWLEQALIPVTARGTEISRVRIPIPHSWAITHGAVILTPECKPDEE
WRAHMLQALPIYQEKLTSMQRLITENMDAKGINAWARLNPEYGEAVYVMVKHRDSTR
LDLNAIGADEIETVPTTEGFIYHRNSNVAWLPTPVKEKGLAVSWLEKLAERGVFPV
IGLDSLSDRHMKLCSWFGIPROQFADAIISORIFGEN"
complement(272800. .273942)
/gene="ECs1349"
CDS
complement(272800. .273942)
/gene="ECs1349"
[note="unknown, similar to hypothetical protein e.g.
[Serratia marcescens] gi|1695869|gb|AAB37123.1| percent
identity 99 in 197 aa, [Deinococcus radiodurans (strain
R1)] gi|7471591|pir||F75301 percent identity 38 in 364"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34772.1"
/db_xref="GI:13360809"
/translation="MSDILEMDIYRRTLSCGTIQVTRDQGEVSLDDLDFDIAERRNPKR
AFLFVSKVLGRHIPVPSVMQAYROLASFSTLGPVLFIGMAETAVGLAGVDE
VRHOEHTSVLTSTRHPVDGTLTCEKEHSHATDLIYLPDDEKRRRVNARTLVL
IDDEATGNTFTNLSALRNTCKLOHLEOVIAVTLTDMSGNALSERTPLPVTSVLSV
GQGWTPLPDAPVDPMPKVNVTSRGEWDIQGKSGRLGLAPADLGHVSVHGER
ILVLGTGERVWEPFLAERLEAAGAQAQFYGSTRSPVAVGYAIESAISFTDNYGLGIP
NFYVNAHQQFDRILVCTTPAESIDTQLLKALAEVAPVVEIYTYE"
complement(273952. .275010)
/gene="ECs1350"
CDS
complement(273952. .275010)
/gene="ECs1350"
[note="unknown, similar to hypothetical proteins e.g.
[Serratia marcescens] gi|1695870|gb|AAB37124.1| percent
identity 99 in 173 aa, [Serratia marcescens]
gi|1695871|gb|AAB37125.1| percent identity 98 in 53 aa,
[Deinococcus radiodurans] gi|7471522|pir||E7530"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34773.1"
/db_xref="GI:13360810"
/translation="MNKKLWFMEGLSSQRDITQGVKSFAQKNNAFIVTFASHRNERNE
ILSVADYSTEPEDPQKRLQFTQETIQTGYIIHHTGRNSQWFEEHRSIAIEPTGATLT
TGATGVDMWTLADEKVTFAQMEQKGLPVPSNRVNTLAEKTLHAAAPTSPGVCK
PVTGIYGMGFWRDSDSPAVMNPHEHRLVSPQOYIAAASAEKPLVLMPLPGP
EFSVDILADKGETLAAVGRKEGAIOYLVNEGSAWELACDCARVMKADGLVNVQTRND
VNGNPIVLEENRPSGCVGYTLHSGVNLPGLENAFKLGLMSEDNVRQSKNFTSPVAV
RSITDVIATPESLSNLIN"
275333. .275914
/gene="ECs1351"

CDS
275333. .275914
/gene="ECs1351"
[note="possible tellurium resistance, similar to Terz
proteins e.g. [Serratia marcescens] gi|6094454|sp|Q52353|
percent identity 98 in 193 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tellurium resistance protein Terz"
/protein_id="BAB34774.1"
/db_xref="GI:13360811"
/translation="MVSLVKQTVSLSKESSALSQHLGWDVPVKKKGLLGLGFGGN
DSIDLDAAGVLMDSGTGKTDITIFRKLKSTCGAVHSGDNLTEGGDDGVIVKLNLSR
LPANVEYLAFATVNSFRGQSFENAFRCRVVDQTKELARYKLTEQSGHTGIVISSLR
RNNGNDFETALGHACRGRTIDDMHSDIVSAVR"
276046. .277071
/gene="ECs1352"
CDS
276046. .277071
/gene="ECs1352"
[note="tellurium resistance, similar to Tera proteins e.g.
[Serratia marcescens] gi|5702379|gb|AAD47285.1|AFI68355_3
percent identity 67 in 385 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tellurium resistance protein Tera"
/protein_id="BAB34775.1"
/db_xref="GI:13360812"
/translation="MVFYQGRNDDGTVSLVSEGOYSTFTVALNRLKPDVQKIAFTVTT
CDGGQTVSGLRNLSDIVEQATGLVSGVSELSGRQEAALILGEFYRNNNDKFRFVQ
GFNGGLKPLAEHFGVNIADPEAPAPTPPVVTPPVTKISLSKVLTKREKPAISLTJK
RDNFGIEIRINLNRHSGSSKGFAGMFGSKGIDLDLGAFLVLODGYKSVIOALGNAGFD
YRDEPVQLGDDORTGDVSDGEMLHNGREWKHIREVLIYAFIYEGVPSWDKTDGVVT
IHWDPDPIETRLTEGENRRTLCIARLNVENCAIKVERINQYFKGQDENDRAFGWGF
RWSAGSK"
277094. .277549
/gene="ECs1353"
CDS
277094. .277549
/gene="ECs1353"
[note="tellurite resistance, similar to TerB proteins e.g.
[Serratia marcescens] gi|950680|gb|AAA86848.1| percent
identity 100 in 151 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tellurium resistance protein TerB"
/protein_id="BAB34776.1"
/db_xref="GI:13360813"
/translation="MSFFDKVGAINSGRDELTRQVGRFNKKFMQGTAVCAARIAS
SDGVSESEKQKMMGFLRSEELKVFDTNEVIEFNKLVSSFDFDVEIGKGETMKYILA
LKQDPEAAQALAVGIATAVAKSDGDFDQDEKLASREITAIAGFDPAEFGL"
277572. .278612
/gene="ECs1354"
CDS
277572. .278612
/gene="ECs1354"
[note="tellurium resistance, similar to TerC proteins e.g.
[Serratia marcescens] gi|6226214|sp|Q52356|TERC_SERMA
percent identity 100 in 346 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tellurium resistance protein TerC"
/protein_id="BAB34777.1"
/db_xref="GI:13360814"
/translation="MVSTHIGFTETVIVFIALSVGAIFIDLPMHRDDKPISLKSAAL
WSFVWVWMAFAGFLYIHGAEVASLFTVGALEKLVSDNLFVMMAIWSFAVDPDR
VHRVLYGLIGIAVFRGFIIVAICTSLSLGYPVEVVFVIAIVAVTAVMLKSGDDDE
IEDYSOHLAYVMKVRFFPIWPKLRGHAFILNOKVEAEALAKPENSQVDTIGRGKAALY
ATPLFLCAVVELSDYMAFDSVPATIAVRSREPLIVYSAMMFAILGLRITLYFVLEALK
QYLVHLEKAVIVLLFFLIAKLGILNATDHIWHHGYSTAAFTTSLYVVLGVLAGLILASVM
FPGKPSSEKGS"
278661. .279239
/gene="ECs1355"

CDS

278661. .279239
 /gene="ECS1355"
 /note="tellurium resistance, similar to terD proteins e.g. [Serratia marcescens] gi|6094448|sp|O52357|TERD_SERMA percent identity 100 in 192 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative tellurium resistance protein TerD"
 /protein_id="BAB34778.1"
 /db_xref="GI:13360815"
 /translation="MSVSLSKGNNVLSKAAPSKNNVLVLGWDARSDTGQDFLDAS AFLASNGKVRGDSDFIYNNLTSSDGSVHTGDNRTGEGDDDES LKIKLDVAPSEV DKLIFFVYTIHDAQARRQSGQVSAFIRLNDNQTEVARYDLTEDASTETAMLFGL YRHNGWKFRVGGGYAGGLASVCAQYGINAS"
 279308. .279883

gene

CDS

/gene="ECS1356"
 /gene="ECS1356"
 /note="possible tellurium resistance, identical to gi|7108482|gb|AAAF36434.1|AF126104_3 TLRB_ECOLI percent identity 100 in 191 aa, similar to TerE proteins e.g. [Serratia marcescens] gi|6094449|sp|O52358|TERE_SERMA percent identity 98 in 191 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative tellurium resistance TerE"
 /protein_id="BAB34779.1"
 /db_xref="GI:13360816"
 /translation="MAVSLVKGNNVSLTKAPTNNVAMVGLGNDARVTDGQGFLLDAS FEAVGEDGKVLSDAHFIFPNKNTSPDGAVEHQDNRGTGEGDDEQVKIDLTKVSDI KVLFFAVYTIHDAEARKQNFQGMVSNFMVYNNNDNGTEIARFDLSEADASTETAMVFGEL YRHGTETKFRVGGQFAGGLSALASQHGNNV"
 279908. .280111

gene

CDS

/gene="ECS1357"
 279908. .280111
 /gene="ECS1357"
 /note="unknown"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="hypothetical protein"
 /protein_id="BAB34780.1"
 /db_xref="GI:13360817"
 /translation="MPGFFLIGDPLSTGEYKGGRKQSYVNSGMNLLCRPRVYLMWL VKASIPCLISTACPSDICYFCDL"
 280305. .280613

gene

CDS

/gene="ECS1358"
 280305. .280613
 /gene="ECS1358"
 /note="tellurium resistance, identical to gi|7108481|gb|AAAF36433.1|AF126104_2 TRLA_ECOLI percent identity 100 in 102 aa, similar to TerF proteins e.g. [Serratia marcescens] gi|7387491|gb|AA86852.2| TREF_SERMA percent identity 94 in 102 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative tellurium resistance protein TerF"
 /protein_id="BAB34781.1"
 /db_xref="GI:13360818"
 /translation="MNLQSQNPLOQSTIRLNLQYPAKSGFKGPDPTCLFMLNAGGK VSGSDSDFIYNNLSSPEGAVRLVTSQQASIEIALDRVPANVSKIAITVVIDGDTIS"
 complement(280747. .280971)

gene

CDS

/gene="ECS1359"
 complement(280747. .280971)
 /gene="ECS1359"
 /note="unknown, GTG start"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental

/product="hypothetical protein"
 /protein_id="BAB34782.1"
 /db_xref="GI:13360819"

/translation="MTFOGNNFMRLAGOVLTITLTWRSHVSGSIPLSFAANTDQAVHH CRTIARHLKVLRSROQEGILSRHFSRGPVR"
 complement(281205. .283295)

gene

CDS

/gene="ECS1360"
 complement(281205. .283295)
 /gene="ECS1360"
 /note="ECS1360"
 /note="adhesin, similar to Iha adhesin [Escherichia coli O157:H7 strain 86-24] gi|7108480|gb|RAAF36432.1|AF126104_1 IHA_ECOLI percent identity 99 in 696 aa, also similar to exogenous ferric siderophore receptor R4 [Escherichia coli strain CFT073] gi|3661500|gb"
 /codon_start=1

/transl_table=11
 /evidence=not_experimental

/product="Iha adhesin"
 /protein_id="BAB34783.1"

/db_xref="GI:13360820"

/translation="MRITTLASVVIPLGFSASSIAAAEDVMIVSASGYEKKLTNAAS SVSVISQELQSSQYHDLAELARSVGEVDVSGTGKGTGLEISIRGPASVYTLILIDG VROGSSDVTPNGFSAMNTGMPPLAAIERIEVIRGPMSTLYGSDAMGGVNIITRKN ADKWLSSYNAGLNQESKNWGNSSQFNWSSGPLVDDSVSLQVRGSGTQROGSSVTSL SDTAATRIPIPTESQNYNLGARDLMDKASBODLVDFMDTTRORYDNRDQGLSGTGGY DRTLRYERNKISAGYDHTFTFTGTMKSYLNNMETENKRELVRSLVRDQKWSVFAEDE LKESNLILNLLTPLGESHVTVGGEFQSSMKDGVVLASTGTFTFKQKWSVFAEDE WHLTDALALTAGSRYEHHQFGGHFSRAYLWVDVADAWTLKGGVTTGYKAPRMGOLH KGISGVSGQKTNLLGNPDLKPEESVSYEAGVYDNPAGLNANTYRGTWDFSNKIVSY SINDNTSNYSNGKARLUHGEFAGTLPSESDVTLSLNTYRSEQDGDGDNKGAPLSY TPHEVNAKLNMQITEVASMLGARYRGKTPRTQNTSSLSAVOKKYDEKGEYLKAW TVVDAGLSKMTDALTLNAAVNLLNKDYSDVSLYSAGKSTLYAGDYFQTGSGSTTGYV IPRNYSMNSLYQF"

gene

CDS

complement(284099. .284221)
 /gene="ECS1361"
 complement(284099. .284221)
 /gene="ECS1361"

/note="unknown, similar to part of hypothetical protein [Shigella flexneri] gi|5880472|gb|AAD54665.1|AF097520_3 percent identity 44 in 40 aa"

/codon_start=1
 /transl_table=11

/evidence=not_experimental
 /product="hypothetical protein"

/protein_id="BAB34784.1"

/db_xref="GI:13360821"

misc_feature

/translation="MSGKTEFVDVICFVYFIIVCVKSFKLIPLKWKAPGFRK"
 284269. .284565
 /note="partial CDS, similar to part (117-213 in 347 aa) of ShiA [Shigella flexneri] gi|5532447|gb|AAD44731.1|AF141323_2 percent identity 97 aa"

gene

CDS

284694. .284966
 /gene="ECS1362"
 284694. .284966
 /gene="ECS1362"
 /note="unknown, similar to C-terminal part of ShiA [Shigella flexneri] gi|5532447|gb|AAD44731.1|AF141323_2 percent identity 49 in 73 aa, TTG start"
 /codon_start=1
 /transl_table=11

/evidence=not_experimental
 /product="hypothetical protein"

/protein_id="BAB34785.1"

/db_xref="GI:13360822"

/translation="MPCNNTYNSVSVLKERSEMFKYAEMLMAEEDKVCCELIFKQODN LILKINVRPRTIIKKYRTLQGVVWMLTLPFVIELVLTIGVNS"
 complement(285988. .286363)
 /note="IS1151, truncated (5' and 3')"

gene

CDS

complement(286087. .286239)
 /gene="ECS1363"
 complement(286087. .286239)
 /gene="ECS1363"

misc_feature

/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34786.1"
/db_xref="GI:13360823"
/translation="MLYDAFTVPGQRP RR CIWIRVGNIELQAI GKVKAPWSSTKYAA
QRKRSE"
complement(286244..286408)
/note="partial CDS, similar to part of IS150 putative
transposase gi|790866|gb|AAA96489.1| percent identity 40
in 52 aa"
gene
complement(286715..286909)
/gene="ECs1364"
CDS
complement(286715..286909)
/gene="ECs1364"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34787.1"
/db_xref="GI:13360824"
/translation="MNENKIKRLEQLQARQREFATKWRGAALPHERLTYFPLEDLIK
NNGKKIRSDRKKHFECEDEK"
gene
286989..287213
CDS
/gene="ECs1365"
286989..287213
/gene="ECs1365"
/note="unknown,TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34788.1"
/db_xref="GI:13360825"
/translation="MELSTKGDEARLYRNGSMALMPVCGSIMASLSIRCCSRKKNFA
SSGDKSIVIRKAMRCSALRYADGSYPF"
gene
287229..287501
CDS
/gene="ECs1366"
287229..287501
/gene="ECs1366"
/note="unknown,GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34789.1"
/db_xref="GI:13360826"
/translation="MSYOPGGLVPLDLSRRGYLSKNASSNRFLGVNGDGLSKDSRCW
CTFSRWRSSMDNAIACSSSLVICTWEAIYRSYSHYGNVLLILOA"
gene
287806..288000
CDS
/gene="ECs1367"
287806..288000
/gene="ECs1367"
/note="unknown,TTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34790.1"
/db_xref="GI:13360827"
/translation="MCHRAHQSCSTRDVALDRHLPCDCEDILKEVIWAFSDFVRNRGV
YDPAIRYPAGNPWYPTVGQF"
gene
complement(288048..288194)
CDS
/gene="ECs1368"
complement(288048..288194)
/gene="ECs1368"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental

gene
complement(288170..288478)
/gene="ECs1369"
CDS
complement(288170..288478)
/gene="ECs1369"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34792.1"
/db_xref="GI:13360829"
/translation="MTNQSGRDNNPLLNGRYOEDEREYRQHOGQGRVPOYQRWHFAH
CQRTGITTRFMLLPPIPRFVATSTALLHFHCSTSLLYQRLRLRPSLMGLGKWPSSLNL
"
gene
288993..290117
CDS
/gene="ECs1370"
288993..290117
/gene="ECs1370"
/note="possible glucosyl-transferase, similar to
glucosyl-transferases e.g. [Salmonella typhi]
gi|7467230|pir|I730292 percent identity 72 in 366 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative glucosyl-transferase"
/protein_id="BAB34793.1"
/db_xref="GI:13360830"
/translation="MRKRILFIGPPLYGLLYPLISLAQAFRVIGHVIVVVISSAGKFANK
AAEAGLVFDAVPGLDSEAGYRHOELRKKSNIIGHFSFSDENADNLIDFAGKWRPD
LIVYPDLPAGPLVAARYIPSVMLAVGEAHTSAHIONLRSLSNAYRRHGVSGPLCD
LAWIDAPPMSILKNAGEPVISMRYPYNGGAVKETWDRSDSRKRLILSLGTVKPK
VDGLELSWMDSANEDADIILQILAINARTGLRKLPSNRLVDWIMPMGVFLNGADGF
IHGGACNTITALYSGIPQIVFGGADRVSNAEIVAKRGCGIIPDKHGLTSDLVNRRL
YDLSRFCSDQAAEMAEQPSAEIAEVLMRKLNKGQL"
gene
complement(290259..290534)
CDS
/gene="ECs1371"
complement(290259..290534)
/gene="ECs1371"
/note="TTG start
probable ferric enterochelin esterase (partial),
similar to C-terminal parts of ferric enterochelin
esterases e.g. [Salmonella enterica] gi|
2738250|gb|AAC46181.1| percent identity 66 in 68 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferric enterochelin esterase"
/protein_id="BAB34794.1"
/db_xref="GI:13360831"
/translation="MESLPENNTSWVSEQLSAPPKDVNIRLGVSLGTTVSHVQRL
HHSLIAAGLESNLTVYTGHDYAWWGAIIDELANYNCRKVSNNSV"
complement(290521..291288)
/note="ISI"
gene
complement(290536..291039)
CDS
complement(290536..291039)
/gene="ECs1372"
/note="transposase, similar to transposases e.g. [Shigella
boydii] gi|2197010|gb|AB61273.1| percent identity 100 in
167 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="transposase"
/protein_id="BAB34795.1"
/db_xref="GI:13360832"
/translation="MSRQCTHYGRWPQHGTSLKKLRPQSVTSRIQPGSDVIVCAEMD
EQMGYVGAKSRQRLWFAYDRIRRTVVAHVFGERTLATLERLLSLSAFEVVMMDTG

repeat_region

WPLYESRLKGLHVISKRYTQRIERHNLRLQHLARLGRKSLSEFSKSVELHDKVIGHY
LNKHKQ
/gene complement(290958. .291233)
/gene="Ecs1373"
CDS complement(290958. .291233)
/gene="Ecs1373"
/note="possible repressor, similar to Insa proteins e.g.
[insertion sequence ISiF] gi|124915|sp|P19767|ISA2_ECOLI
percent identity 98 in 91 aa,GTG start"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="Insa protein of insertion sequence IS1"
/protein_id="BAB34796.1"
/db_xref="GI:13360833"
/translation="MASISIRCPSCATEGVVRNGKSTAGHQRYLCSPCRKTWQLQFT
YFASQPKHQKIIDMAMNGVGCASARIMGVGLNTVLRHLKNSGRSR"
complement(291471. .291929)
/gene="Ecs1374"
CDS complement(291471. .291929)
/gene="Ecs1374"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative membrane protein"
/protein_id="BAB34797.1"
/db_xref="GI:13360834"
/translation="MKIYWRKRIPELSALPPSLRKKNTDAYNAASSHTIYEWIGAV
SFISMLSRVDFLLPAODTPGDIIISLCVVCPSILLWFQSVYMKRYRHLVR
GKETIISBLREADTREYELMRPVRFFSIVFLVLLGCIHSLITTK"
291994. .292185
/gene="Ecs1375"
CDS 291994. .292185
/gene="Ecs1375"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34798.1"
/db_xref="GI:13360835"
/translation="MTTSPFILEFHDNRDNRHLYQLIVSYLITLWFWNQVIDLMKR
VGLWYCVVLFGICSYIN"
292387. .292896
/gene="Ecs1376"
CDS 292387. .292896
/gene="Ecs1376"
/note="unknown,GTG start"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34799.1"
/db_xref="GI:13360836"
/translation="MQITELISEGEIRRFVQQAVDHPNLLAFHFTLYSAGGIYGO
RQGVSPHRVHERITEHNHTVSPAPVLRLEQHEGAQIRCLLLSQTSTCHP
RIGVMADECAQLVDLLQQTWSVISAGGQCRVRCFRVARGSGGYVALKTAQVSFM
SOVIATIR"
292985. .293608
/gene="Ecs1377"
CDS 292985. .293608
/gene="Ecs1377"
/note="unknown,weakly similar to sensor-regulator protein
HutT [Rhodobacter capsulatus] gi|1075537|pir|IA49938
percent identity 33 in 97 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34800.1"
/db_xref="GI:13360837"
/translation="MRTRIDYLADKYCFERNESPLRRQWDVLEECRQTAGPEER

LRIALLNVYVTSPELPRLLLTRPOLIAALREEWGISQKNVFNDRKRCVVSLSKA
SLSGVPTIYRHLSHRIRRVVNGENTISLPYQQVAREKAPRERLKYALEAGLLTALD
GLFRGQSRIAADVLRLRKAGMPVTTTVEVDNLTGTRKVPAYH"
293704. .293937
/gene="Ecs1378"
CDS 293704. .293937
/gene="Ecs1378"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34801.1"
/db_xref="GI:13360838"
/translation="MARSYDIIINLSLEHELNEMLAERGYAGLVDRNRRLAEVVTRKL
QDSFYINVSQDALNTAYSEHPWFSGLVSGDEN"
293948. .294163
/gene="Ecs1379"
CDS 293948. .294163
/gene="Ecs1379"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34802.1"
/db_xref="GI:13360839"
/translation="MLLGGSVIAGRDFVLCFFPGFGEFSQKVHVRSHAGMRTAVLH
KHHPFLNTQGDKNMIVFVSLIDFR"
complement(294733. .296042)
/note="IS629-03"
/gene="Ecs1380"
CDS complement(294775. .295662)
/gene="Ecs1380"
/note="transposase, similar to transposases e.g. [insertion
sequenceIS629] gi|7443863|pir|T00315 percent identity 98
in 295 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative transposase OrfB protein of insertion
sequence IS629"
/protein_id="BAB34803.1"
/db_xref="GI:13360840"
/translation="MPLLDKLEHQYGVPCSELHIAPISTYVHCQQORHHPDKRSARA
QRDDLKREIQRYVDENHOYGVYKVMRQLLRGIRVACTVAELMVGMLAGVLRCK
KVRTYSRKAVSAGDRVNRQFVAERPQLWVADFTYVSTWQGFYVAFIDVFAAGCIV
GWRVSSMETTFVDLALEQALWARRPSGTVHHSDKGQYISLAYTERLKEAKLLASTG
STGDSYDNAMAESINGLYKAEVIHRKSKNRAEVELATLTWVDWYNNRRLGLRLGHIP
PAEAERAYVASIRNDLIA"
complement(295662. .295988)
/gene="Ecs1381"
CDS complement(295662. .295988)
/gene="Ecs1381"
/note="transposase, similar to hypothetical proteins e.g.
[Escherichia coli plasmid p0157 insertion sequence IS629]
gi|7444868|pir|T00241 percent identity 96 in 108 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative transposase OrfA protein of insertion
sequence IS629"
/protein_id="BAB34804.1"
/db_xref="GI:13360841"
/translation="MTKNTRFSEVRQAVRMVLESGEYDSQWAAICSTAPKIGCTPP
ETLRVWRQHERDTSGDGLTTAERQRLKELENRELRSNDILRQASAYFAKEEF
DRLMKK"
296481. .297215
/gene="Ecs1382"
CDS 296481. .297215
/gene="Ecs1382"
/note="HecB-like protein,

N-terminal half is similar to N-terminal parts of hemolysin activation protein HecB [Neisseria meningitidis MC58] gi|7227016|gb|AAFA42103.1| percent identity 34 in 181 aa"

```

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HecB-like protein"
/protein_id="BAB34805.1"
/db_xref="GI:13360842"
/translation="NSTLQNLVDHGVYVTRVLPASQDLKSGILRLVLPGVVRVRL
TPDSDYIQLYSFPAHEGSLDLRDIQQGLDLSNRQCGHTELNATSGNLSTONAQ
LSADTLRTAGQFSNGGTINADTLQISAQSLNRKGLSIGTGDFSLSLPQSDVN
REGLLAANGAVRLDALSDNRKGVQAQSPSLQKSPPTFLKPFVAGVCAALLAVSA
IPGWQFLQPSPEEQHFTWNGCKKQ"

```

gene

CDS

```

297456..297971
/gene="ECs1383"
297456..297971
/gene="ECs1383"
/translation="unknown"
/codon_start=1

```

/transl_table=11

```

/evidence=not_experimental
/product="hypothetical protein"

```

/protein_id="BAB34806.1"

/db_xref="GI:13360843"

```

/translation="MVTICVDGDKTHKITDWTWAGNDREVMLTCHFRSGRKYTRP
LSVCOITPTVNLNRVFLERKGNVTSRAELVIYGDKVAAYVYREGERPIMKTTGLD
FQCSAFTEHAVENYLCRVANERIFYARGNRENIDENILRQIKKIVPPDPTALHAYS
GQSKNVIRHGVT"

```

gene

CDS

```

298199..300136
/gene="ECs1384"
298199..300136
/gene="ECs1384"

```

/note="unknown, weakly similar to part of tRNA-splicing endonuclease positive effector [fission yeast] gi|7493527|pir||T40065 percent identity 22 in 531 aa, also weakly similar to hypothetical proteins e.g. [Aquifex aeolicus] gi|7514764|pir||D70476 percent id"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="hypothetical protein"

/protein_id="BAB34807.1"

/db_xref="GI:13360844"

```

/translation="MARLGSTENROQFFSTLSRSEEVLPDPSANADDDVLQVKKH
LNAINQVSLKRAEINELNIEKYLQWQSONLRPELFSHKTRFSQKTTDLMAIHY
LSDRIGFRNRIDLLNFMILKVKPLMIPERRLALFTSLQSYBKNIKEKOISLNEY
EEAFKSDFKILLGRITSMYLYLKOHLRNRVYSTRSSAETRYDEFDFIKRFTIIG
SSTHSIINSIGKALLDYVIIDEASQODIVPGILGICARNVIVVGDRLQLPHVPVLL
PNSPSPAAYYNCYKSLSDSVCMFLRNMPVPTLLKEHYRCHPKTIOFCNKOFYNAL
IPLTVDSGEASLSVITAKGNHTRFNSNLRLESLGHHWDESSRGYIAPYNAQVNL
AKVLPADFVSTVKFGRECEDEIVFTVLDDKRSQHSRNIAFVDPNELYNAVSR
ARKFTLVGNDVFRFHAGHTAALIRYKYADGGEIFESPVISAPDLLYSYDKSLE
RLNRLNDSHFKEQIVACILRDLISQDSYRSMFHQSIALNQLVLLERGDFTFRE
QLFMNRASCDVYVYKVGKPTPLGVIEDGGYHLTSVQAERDELKNSILKKGCLPLLR
LRTIDSDIEGKLGAFLSGLTG"

```

gene

CDS

```

300180..300320
/gene="ECs1385"
300180..300320
/gene="ECs1385"
/translation="unknown, TTG start"

```

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="hypothetical protein"

/protein_id="BAB34808.1"

/db_xref="GI:13360845"

```

/translation="MVWCHTQLYEPHGHLRLFCGYINVRIPDTYKKNKVCYCKLTGVS
VP"

```

gene

CDS

```

300637..301869
/gene="ECs1386"
300637..301869

```

/gene="ECs1386"
/note="unknown, similar to hypothetical proteins e.g. YbdN [Escherichia coli] gi|3024984|sp|P77216|YBDN_ECOLI percent identity 58 in 396 aa"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="hypothetical protein"

/protein_id="BAB34809.1"

/db_xref="GI:13360846"

```

/translation="MSDTLLTEKILITGENVLAARIIEWIFETPPSVCLSFSGGKDS
TVLFHLVAEVARRRKHFSVLFIDWEAQRYCTLEHQKRMHYHDVTEYHVALPLT
TVNGVSQPEWICWEPGWVYRQPEEAITDWTTFPFRIAMTEEFYPAESSFAG
NRCGVAVUTGVRADESLNREMGVLSQRKYADKPWTTSPEGYTYTPIYDKMRK
DIWYINARACAIYNPLYDLMYRADYPLRNMRYCEPFGPEQRKGLWLYHVLPEYARM
CEVSGAASGALYNESGAYFALKRKRIKSKPAHHTWRSYAMFLDDVMPERTASHYRNKI
AYVLRWYQTRGFPDDIPDQENDLGSRDIPSNRRICKTLIKNDFWCRLPSESPNKRPH
YERYLQMKRKKEWGL"

```

gene

CDS

```

301854..302492
/gene="ECs1387"
301854..302492
/gene="ECs1387"

```

/note="unknown, similar to hypothetical protein YbdM

[Escherichia coli] gi|3024983|sp|P77174|YBDM_ECOLI percent identity 56 in 212 aa"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="hypothetical protein"

/protein_id="BAB34810.1"

/db_xref="GI:13360847"

```

/translation="MGDSVTPETIIGNMIROYFSQERPEETIOALNHLRRLVHEVSPF
AHEPVDCLVWKADEVANDYNPYMAPEGKLLKQSLKDGFTOPVVYSEKKNHYLV
VDGFHROVLGREGATGKRLKGLPVVCINPERKGOASRIAATIRNRARGKHQITMS
DIVRLSLRGWTNERIGTELGMDQDEVLRKQISGLTELFQEDFSPANTVR"

```

gene

CDS

```

302571..302840
/gene="ECs1388"
302571..302840
/gene="ECs1388"

```

/note="possible transcriptional regulator similar to PerC [Bfpw] [Escherichia coli] gi|1172431|sp|P43475|PERC_ECOLI percent identity 25 in 83 aa"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="putative transcriptional regulator"

/protein_id="BAB34811.1"

/db_xref="GI:13360848"

```

/translation="MVAQKLEAAGCWRRASDRWLFMVGNVECTAQOREWLLLRNYCL
A0ISSPPLPTELDISEVAKAADATLRRMGIAATPSGEVFRKGPVC"

```

302861..303505

/gene="ECs1389"

302861..303505

/gene="ECs1389"

/note="unknown, weakly similar to exopolyphosphatase

[Pseudomonas aeruginosa] gi|4200042|dbj|BAA74460.1| percent identity 32 in 56 aa"

/codon_start=1

/transl_table=11

/evidence=not_experimental

/product="hypothetical protein"

/protein_id="BAB34812.1"

/db_xref="GI:13360849"

```

/translation="MNKQGMANSISMTWQLLCDEHGFTRHYNOLKKFSPETLREIIA
ETASCHPSTSVLLRNKWLTPPEDILEQITREYERRIQNCPPFRSEKAEASWLNELSF
VIAPIRVAPEQTEQVSVLQIAEQERWELILTNDGYSWHCALYDLFLHLVLRNM
EDQPEHQKQRIKTIKFIYEPDARLVAEEIKFRIQSLYDDEQKRALSELVNDFTSKE"

```

gene

CDS

```

303553..303708
/gene="ECs1390"
303553..303708
/gene="ECs1390"
/translation="unknown"
/codon_start=1

```

```
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34813.1"
/db_xref="GI:13360850"
/translation="MKVNGVTSKYEGRCTFQLYRYKTVVLVHPKGYLAGEMIRHQA
LTSSQD"
gene
303821..304321
/gene="ECs1391"
CDS
303821..304321
/gene="ECs1391"
/note="unknown,N-terminal part is similar to BfpM
[Escherichia coli] gi|847983|gb|AAC44052.1| BpM_ECOLI
percent identity 52 in 113 aa,N-terminal part is also
similar to putative transposase [Vibrio cholerae]
gi|7467523|pir||T09435 percent identity 55 in 68"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="BfpM-like protein"
/protein_id="BAB34814.1"
/db_xref="GI:13360851"
/translation="MNETIQRLQWKLYETSGDAGVCRRCGISRPTLRKMKVAQISG
SGHRWGGKPPQSPAFTINENHLRSLSDSGVTFSAKSGKTYSKRVKTTSENLGNS
NNQSPPEPESGLNRPGPGECFICELRLPDHFRWKNKLFLLLPEYGPAPFAYDC
YTSPP"
repeat_region
complement(304161..304419)
/note="IS629-04, truncated (5')"
304269..304490
/gene="ECs1392"
304269..304490
/gene="ECs1392"
/note="ECs1392"
/note="unknown,weakly similar to hypothetical protein
[Pseudomonas syringae] gi|1196744|gb|AA88435.1| percent
identity 34 in 50 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34815.1"
/db_xref="GI:13360852"
/translation="MAQPFQSSIVIPVHPREGQGFHCTVFPALTYYVYLRFKTIDA
LRHRVYSKRKLAVCSHQKIGNLDAHL"
repeat_region
304420..305000
/note="IS682-2, truncated (3')"
304505..304930
/gene="ECs1393"
304505..304930
/gene="ECs1393"
/note="ECs1393"
/note="unknown,similar to hypothetical proteins e.g. orf29
[Escherichia coli] gi|6009405|dbj|BAA84864.1| percent
identity 40 in 131 aa, L0013 [Escherichia coli]
gi|13414891|gb|AAC31492.1| percent identity 38 in 130 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34816.1"
/db_xref="GI:13360853"
/translation="MELODWKREPKRKNYSNEFKLRMWELASOPGASVARIAREHDIND
NLLFKWLQWKEGRISRLQVTTSSDTGVELLPVEITPDEKEPVAATAPSLSTSQ
TSVSAGSKVEPRHGMNMLENPSPELLILLIRELIGRGR"
repeat_region
305027..305075
/note="IS2-1, truncated (3')"
complement(305230..305412)
/gene="ECs1394"
305027..305075
/gene="ECs1394"
/note="ECs1394"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
```

```
gene
CDS
/protein_id="BAB34817.1"
/db_xref="GI:13360854"
/translation="MYNCGESYSEYDLSDDENYDDYRERPCPGEVFGVDNTEDN
DEAASDYALKINENS"
gene
305741..306613
/gene="ECs1395"
305741..306613
/gene="ECs1395"
/note="unknown,similar to hypothetical proteins e.g. yfjP
protein [Escherichia coli] gi|7449539|pir||B65042 percent
identity 49 in 289 aa, yeeP protein [Escherichia coli]
gi|2495624|sp|P76359|YEEP_ECOLI percent identity 95 in 183
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34818.1"
/db_xref="GI:13360855"
/translation="MNPSDAIEAIEKPLSLPYSLSHILEHLRKLKLTSHPEVIGIMSK
SGAGKSLCNALFQGEVTPVSDVHAGTREVRRFLSGHSHSVITDLPVGVGESRDRA
EYALYRDLPELDLVLWLKADDRALSDVDEYFWRHILHRGHOOVLFVVTQADKTEPC
HEWDMAGIOPSPAQEQNIKEKTDVAFELRPVHPVAVSARTGWELDTLVSALMTALP
DHAASPLMRLQDELCTESVWGQAREQFTGAVDRIFDTVESVCVASVARTVLRAVRDT
VVSARAVWNWIFF"
gene
306985..309834
/gene="ECs1396"
306985..309834
/gene="ECs1396"
/note="also similar to Adhesin AIDA-I precursor
[Escherichia coli plasmid pIB6]
gi|5437881|sp|Q03155|AIDA_ECOLI percent identity 23 in 678
aa
also similar to glycoprotein [Escherichia coli strain
H10407] gi|5305639|gb|AAD41751.1| percent identity 34 in
608 aa,
probable adhesin,
similar to outer membrane fluffing protein [Escherichia
coli] gi|7466262|pir||G64964 percent identity 68 in 927
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Aida-I adhesin-like protein"
/protein_id="BAB34819.1"
/db_xref="GI:13360856"
/translation="MKRHLNTSYRLVMNHTGLVVASLARSKRKAGVAVALSIAA
VTSVPALADKVVOAGETVNDSTLTNHDNQIVFGTANGMTISTGLGELGPDSENTGGQ
WLONGGIAGNTVTVTNGRQVILEGGTASDVIIDGGGQSLGLAVAVNTLNNRGEQWVH
EGVATGTTIINRDGYQSVKSGGLATGTTIITGAEGGPDSDSYTGQVQGTAEKSTIN
KNGRQITLFSGLARDTLIYAGGDSVHGRLNLTNGQYVYHRDGLALNTVINEGGW
QVVKAGGAAGNTIINQNGELRVHAGGEATVNTGTGALVTAAVTICTNRLGNFTV
ENGKADGVVLESGRLDVLHSAQNTLVDDGTLAVSAGKATSVTITSGGALIADS
GATVEGTASGRFSDGTSGQASGLLENGSGFTVNAQGAQNTVVGHRGTLTLLAAGS
SLSGRTQLSKGASMLVNGDVSTGDIIVNAGEIFDNQITTPNAALSRAVAKNSPVTFFH
KLTTNLTQGGTINNRVLNDSNQDLVINGQATGTWLAFTNVGNSNLGVATGTG
QGITRVDAQNATTEEGAFALSPLOQAFNLTNRDSDDEWYLRSENYRAEVPLYT
SMLTQAMDYDRILAGSRGTGVNGENSVRLSIOGHLGHNNGGIARGATPESSGS
YCFVRLGDLRLTEVAGMSLTGVYGAAGHSVDYKDDGSRACVTRVDDAGSLGGYLN
LYHTSSGLWADIVAQGRSLRSMKASDNDFFARGMGWLGSLETGTFPLTDNLMLEPQ
LVYTWQGLSDDGDNAGYVKEFGHSAQHRAGFRGLSHNDTFFEGTSSRDTLRDSA
KHSVELPNWVWPQSVITRFSRGMDSMGTAAGSNMTFSPSRNGTSLDLOAGLEAR
IRENITLGVQAGYAHSVSGSSAEGYNGQATLNMTF"
gene
309945..312329
/gene="ECs1397"
309945..312329
/gene="ECs1397"
/note="unknown,similar to hypothetical protein yJDA
[Escherichia coli] gi|731985|sp|P16694|YJDA_ECOLI percent
identity 32 in 793 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
```

```

/product="hypothetical protein"
/protein_id="BAB34820.1"
/db_xref="GI:13360857"
/translation="MHEKNIALLCDEADRLLOLNINLRQMWDPDVLSPDKNENGLL
FDRKALKRIEELEGQIKTAREEMVLAVGTMKAGKSTTINAIVGOEILPNERNPT
SVPTLLRHVPCKTEPVLHLEHIQPVNLLITLOEKLATPAGQVQASLOQTGOTRELL
DIUADDDWUKNEYHGEDEIFTGLASLNDLVRLEAAAGTEPFDFEYAEVOKPLVIDVAF
SHLVGDACQGTLLTIDTPGNEAGQPMQEMMRDQLOKASAVAYMDYTMQMSKADE
EVKELNALADVAGRLFLVFNKFDKDRNGDGADAVROKVPAMLNSDVLPAASRVYPG
SSRQAYANRALHAGRLKALPVDEAVDDFIREAFGPMVEEDDKDSTKVKKAKEL
WNLSLDQITEVIOSSHSPAAALVDSAAKLMQNAENVSEYLSLRHQLOQSIOSL
QAHITSLLEADQIEECQNOVTDVRNAMEDINTKGELTKVCASDEEELNDYFRSG
KAREQOMLEENSAQPRERNAFPHDIFDGTGNHDMRDFDPSPEIKFSDRAALE
LMTQISTVLSHREAEQPRPELEKIVSGIETGFRGTALYATENAFINARLEDEG
FTVKISFPVSOLOQLAVKINLSALMEERTETVTRRRQSGVGMVTCRWFGTSDLGW
ENDEOVSRSVININKAREVMSLTRYFCELOASTEODINOPVROEIDAFFCAFREK
VEOLRNTLQSSSEDKRDOQOERLTCRLQALNERVELITDSKALREELETML"
312326..312331
/gene="Ecs1398"
CDS
312326..312331
/gene="Ecs1398"
/notes="unknown, similar to hypothetical protein YicZ
[Escherichia coli] gi|731984|sp|P39267|YJCZ_ECOLI percent
identity 30 in 278 aa,67G start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34821.1"
/db_xref="GI:13360858"
/translation="MTSPFTQQADNRVCOVLTCLPKEKFVDFVDFANGIDVAQEHRTAG
ERTFFRLKEGLTGEGAARQNALNASLAQGEASLRWLTEMTTSLATTNVAITRVNDR
VSSLVSDTLARLAHYSADTREOLLTADQVHHKLNHLEEKLHRVDQVQRAQLHQEQIFS
WWSAGVAFSPAGRCYVVALEELRWGAFGDVIRQGETGOVNOQLDILRHKALTQMAQE
SGSASVRLNTLDWLGQGREQADNEWHDAINLWGDWCSEEQHPVWISTTQAAEHLPV
RMPRLCSARLSSESMVDEIFQKGA"
313228..314298
/gene="Ecs1399"
CDS
313228..314298
/gene="Ecs1399"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34822.1"
/db_xref="GI:13360859"
/translation="MSTEMKTLGLVSGGGANGAYQAGVVKALAECCQTQISMVSGTSIG
AFNGALTAASPOLSEAAVLEALWEHLGNQVLSVNRVYFSLKKLFQAMNLCQIPG
RAGALLTLRLHISTINGFDNLMAQPLSDLEPLTALMDHYLDTDALADGLPLYVSLYP
TEGGMODIIDCIRAEILGVGTTKNAVFOHQSLPRGOQKEALLASALPLLFRPREVOG
TMFGDGMGSGWRNMGNTPVTPLVDAGCNMVIYTHLSDGSLWDROAFPPTTILEIRPR
KRLKYAGDGSNGSGLSFTSAHTDANRQOQYEDTMLAMEHIRKPLAARQALTRSEAVL
QKSLDITERADLALRNAMARIK"
complement(314403..314567)
/gene="Ecs1400"
complement(314403..314567)
/gene="Ecs1400"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34823.1"
/db_xref="GI:13360860"
/translation="MGEMNEVCLRSFSLANHVLSCLIPASPVOLTFSIADNPPEICSL
VVFVSSPKIK"
314638..315456
/gene="Ecs1401"
CDS
314638..315456
/gene="Ecs1401"
/notes="unknown, similar to hypothetical proteins e.g. YfjQ
[Escherichia coli] gi|1723629|sp|P52132|YFJQ_ECOLI percent
identity 30 in 278 aa,67G start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34824.1"
/db_xref="GI:13360861"
/translation="MRLASRFYANOIRRDRLPLTHEELMRHVPSIFGCENRHTSRSEHY
AYPTITVLENLOQEGFQPFACQTRVDRSREYTKHMLRLRAGQITQGHVPEIIL
LNSHDSSTQMLPGIFRAICTNGLVCGSLGEURVHGRNVDRVREGAYEVGVGVD
RIEKKDAMOSLVLPAPRALQAALTRYGDEHVPVTADILTPRRREDYDQDWS
AYOTIQENMLKGGISGRSAKGRITRAIHSDITDKLNRLALWMAETLLESMR"
315547..316032
/gene="Ecs1402"
CDS
315547..316032
/gene="Ecs1402"
/notes="unknown, similar to hypothetical proteins e.g. Yafk
[Escherichia coli] gi|2495486|sp|P75676|YAFK_ECOLI percent
identity 71 in 144 aa, YfjX [Escherichia coli]
gi|1723636|sp|P52139|YFJX_ECOLI percent identity 75 in 137
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34825.1"
/db_xref="GI:13360862"
/translation="MTTVSHNSTPVSVTTASGNPNPQLVATVPDQQRISFWPQHF
GLTPQWTLPRVFGIWMDRCLCEDYCGIWNLYTLNNGGAFAPEPDDDDDETWVLFNA
MNGNRAEMSPFAAGIAACLMYSHHACRTECYAMTVHYRLOYALQHPCECSA1MR1I
D"
316048..316524
/gene="Ecs1403"
CDS
316048..316524
/gene="Ecs1403"
/notes="probable DNA repair protein (RadC family), similar
to putative RadC family proteins e.g. YkfG [Escherichia
coli] gi|3025218|sp|Q47685|YKFQ_ECOLI percent identity 81
in 158 aa, Yees [Escherichia coli]
gi|3025155|sp|P76362|YEES_ECOLI percent identity 98"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative DNA repair protein"
/protein_id="BAB34826.1"
/db_xref="GI:13360863"
/translation="MQOLSFLPGMTPGERSLILRALQTLDRHLHPEGVAFSTRAAR
EWILANMAGLEREEFRVLYLNQNQLIAGETLFTGTINRTEVHPREVIRKALYHNAAA
VVLAHNHPGCEVTPSKADRLITERLVQALGLVDIRVPDHLIVGNGQVFSFAEHGLL"
316587..316808
/gene="Ecs1404"
CDS
316587..316808
/gene="Ecs1404"
/notes="unknown, similar to hypothetical protein YeeT
[Escherichia coli] gi|3025156|sp|P76363|YEET_ECOLI percent
identity 97 in 73 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34827.1"
/db_xref="GI:13360864"
/translation="MRLITRGEAMRIHQHPASRLFPFCTGKYRWHGSAEAYTGREVQ
DIPCVLAVFAERKDSFGPYVRLMSVTLN"
316971..317345
/gene="Ecs1405"
CDS
316971..317345
/gene="Ecs1405"
/notes="unknown, similar to hypothetical proteins e.g. YeeU
[Escherichia coli] gi|3025157|sp|P76364|YEEU_ECOLI percent
identity 89 in 118 aa, YfjZ [Escherichia coli]
```

```

/product="hypothetical protein"
/protein_id="BAB34820.1"
/db_xref="GI:13360857"
/translation="MHEKNIALLCDEADRLLOLNINLRQMWDPDVLSPDKNENGLL
FDRKALKRIEELEGQIKTAREEMVLAVGTMKAGKSTTINAIVGOEILPNERNPT
SVPTLLRHVPCKTEPVLHLEHIQPVNLLITLOEKLATPAGQVQASLOQTGOTRELL
DIUADDDWUKNEYHGEDEIFTGLASLNDLVRLEAAAGTEPFDFEYAEVOKPLVIDVAF
SHLVGDACQGTLLTIDTPGNEAGQPMQEMMRDQLOKASAVAYMDYTMQMSKADE
EVKELNALADVAGRLFLVFNKFDKDRNGDGADAVROKVPAMLNSDVLPAASRVYPG
SSRQAYANRALHAGRLKALPVDEAVDDFIREAFGPMVEEDDKDSTKVKKAKEL
WNLSLDQITEVIOSSHSPAAALVDSAAKLMQNAENVSEYLSLRHQLOQSIOSL
QAHITSLLEADQIEECQNOVTDVRNAMEDINTKGELTKVCASDEEELNDYFRSG
KAREQOMLEENSAQPRERNAFPHDIFDGTGNHDMRDFDPSPEIKFSDRAALE
LMTQISTVLSHREAEQPRPELEKIVSGIETGFRGTALYATENAFINARLEDEG
FTVKISFPVSOLOQLAVKINLSALMEERTETVTRRRQSGVGMVTCRWFGTSDLGW
ENDEOVSRSVININKAREVMSLTRYFCELOASTEODINOPVROEIDAFFCAFREK
VEOLRNTLQSSSEDKRDOQOERLTCRLQALNERVELITDSKALREELETML"
312326..312331
/gene="Ecs1398"
CDS
312326..312331
/gene="Ecs1398"
/notes="unknown, similar to hypothetical protein YicZ
[Escherichia coli] gi|731984|sp|P39267|YJCZ_ECOLI percent
identity 30 in 278 aa,67G start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34821.1"
/db_xref="GI:13360858"
/translation="MTSPFTQQADNRVCOVLTCLPKEKFVDFVDFANGIDVAQEHRTAG
ERTFFRLKEGLTGEGAARQNALNASLAQGEASLRWLTEMTTSLATTNVAITRVNDR
VSSLVSDTLARLAHYSADTREOLLTADQVHHKLNHLEEKLHRVDQVQRAQLHQEQIFS
WWSAGVAFSPAGRCYVVALEELRWGAFGDVIRQGETGOVNOQLDILRHKALTQMAQE
SGSASVRLNTLDWLGQGREQADNEWHDAINLWGDWCSEEQHPVWISTTQAAEHLPV
RMPRLCSARLSSESMVDEIFQKGA"
313228..314298
/gene="Ecs1399"
CDS
313228..314298
/gene="Ecs1399"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34822.1"
/db_xref="GI:13360859"
/translation="MSTEMKTLGLVSGGGANGAYQAGVVKALAECCQTQISMVSGTSIG
AFNGALTAASPOLSEAAVLEALWEHLGNQVLSVNRVYFSLKKLFQAMNLCQIPG
RAGALLTLRLHISTINGFDNLMAQPLSDLEPLTALMDHYLDTDALADGLPLYVSLYP
TEGGMODIIDCIRAEILGVGTTKNAVFOHQSLPRGOQKEALLASALPLLFRPREVOG
TMFGDGMGSGWRNMGNTPVTPLVDAGCNMVIYTHLSDGSLWDROAFPPTTILEIRPR
KRLKYAGDGSNGSGLSFTSAHTDANRQOQYEDTMLAMEHIRKPLAARQALTRSEAVL
QKSLDITERADLALRNAMARIK"
complement(314403..314567)
/gene="Ecs1400"
complement(314403..314567)
/gene="Ecs1400"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34823.1"
/db_xref="GI:13360860"
/translation="MGEMNEVCLRSFSLANHVLSCLIPASPVOLTFSIADNPPEICSL
VVFVSSPKIK"
314638..315456
/gene="Ecs1401"
CDS
314638..315456
/gene="Ecs1401"
/notes="unknown, similar to hypothetical proteins e.g. YfjQ
[Escherichia coli] gi|1723629|sp|P52132|YFJQ_ECOLI percent
identity 30 in 278 aa,67G start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34824.1"
/db_xref="GI:13360861"
/translation="MRLASRFYANOIRRDRLPLTHEELMRHVPSIFGCENRHTSRSEHY
AYPTITVLENLOQEGFQPFACQTRVDRSREYTKHMLRLRAGQITQGHVPEIIL
LNSHDSSTQMLPGIFRAICTNGLVCGSLGEURVHGRNVDRVREGAYEVGVGVD
RIEKKDAMOSLVLPAPRALQAALTRYGDEHVPVTADILTPRRREDYDQDWS
AYOTIQENMLKGGISGRSAKGRITRAIHSDITDKLNRLALWMAETLLESMR"
315547..316032
/gene="Ecs1402"
CDS
315547..316032
/gene="Ecs1402"
/notes="unknown, similar to hypothetical proteins e.g. Yafk
[Escherichia coli] gi|2495486|sp|P75676|YAFK_ECOLI percent
identity 71 in 144 aa, YfjX [Escherichia coli]
gi|1723636|sp|P52139|YFJX_ECOLI percent identity 75 in 137
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34825.1"
/db_xref="GI:13360862"
/translation="MTTVSHNSTPVSVTTASGNPNPQLVATVPDQQRISFWPQHF
GLTPQWTLPRVFGIWMDRCLCEDYCGIWNLYTLNNGGAFAPEPDDDDDETWVLFNA
MNGNRAEMSPFAAGIAACLMYSHHACRTECYAMTVHYRLOYALQHPCECSA1MR1I
D"
316048..316524
/gene="Ecs1403"
CDS
316048..316524
/gene="Ecs1403"
/notes="probable DNA repair protein (RadC family), similar
to putative RadC family proteins e.g. YkfG [Escherichia
coli] gi|3025218|sp|Q47685|YKFQ_ECOLI percent identity 81
in 158 aa, Yees [Escherichia coli]
gi|3025155|sp|P76362|YEES_ECOLI percent identity 98"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative DNA repair protein"
/protein_id="BAB34826.1"
/db_xref="GI:13360863"
/translation="MQOLSFLPGMTPGERSLILRALQTLDRHLHPEGVAFSTRAAR
EWILANMAGLEREEFRVLYLNQNQLIAGETLFTGTINRTEVHPREVIRKALYHNAAA
VVLAHNHPGCEVTPSKADRLITERLVQALGLVDIRVPDHLIVGNGQVFSFAEHGLL"
316587..316808
/gene="Ecs1404"
CDS
316587..316808
/gene="Ecs1404"
/notes="unknown, similar to hypothetical protein YeeT
[Escherichia coli] gi|3025156|sp|P76363|YEET_ECOLI percent
identity 97 in 73 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34827.1"
/db_xref="GI:13360864"
/translation="MRLITRGEAMRIHQHPASRLFPFCTGKYRWHGSAEAYTGREVQ
DIPCVLAVFAERKDSFGPYVRLMSVTLN"
316971..317345
/gene="Ecs1405"
CDS
316971..317345
/gene="Ecs1405"
/notes="unknown, similar to hypothetical proteins e.g. YeeU
[Escherichia coli] gi|3025157|sp|P76364|YEEU_ECOLI percent
identity 89 in 118 aa, YfjZ [Escherichia coli]
```

```

gil1723638|sp|P52141|YFJZ_ECOLI percent identity 66 in 98
aa,GTG_start=1
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34828.1"
/db_xref="GI:13360865"
/translation="MSDKLSGITHPPDNDHDPWGLPCTVRPCFGARLVOEGNRLHYL
ADRAIGRFSADAYHLDQAFPLMKQLEMLTSGELNPRYQHTVTLNAKGLTCEAD
TLGSCGVYLAIVPTPAPATTS"
317392..317766
/gene="ECs1406"
317392..317766
/gene="ECs1406"
/translation="unknown,similar to hypothetical protein e.g. L0007
[Escherichia coli] gi|3414875|gb|AAC31486.1| percent
identity 93 in 124 aa, Yeev [Escherichia coli]
gi|3025158|sp|P76365|YEEV_ECOLI percent identity 87 in 124
aa, YKI [Escherichia coli] gi|3025213|s"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34829.1"
/db_xref="GI:13360866"
/translation="MKTLSDTHREVSRCPSPVITWTLIRLIDQHYGLTNDTPPA
DERVIEQIHAGISLDAVNFLVEKVALVTDQPGFSTCPSQLNSIDILRRARTG
LMTRDNYRTVNNITLGKYPEAK"
317763..318254
/gene="ECs1407"
317763..318254
/gene="ECs1407"
/translation="unknown,similar to hypothetical proteins e.g. L0008
[Escherichia coli] gi|3414876|gb|AAC31487.1| percent
identity 94 in 163 aa, Yeev [Escherichia coli]
gi|3025160|sp|P76366|YEEW_ECOLI percent identity 65 in 55
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34830.1"
/db_xref="GI:13360867"
/translation="MMKALITLEADSVNVAQNGRIVVDVGIEIAELINMVCDNGY
SLRVDESOTSAECTPPFATLTGIRGSTAHITETDNANLYSLSHOTSDFGSEWIHF
TNGYLLRTDAWSPVPLRLKRLGSLKTFRLVVTLLIRRYGVSLIHLDAASACLPLPT
FDW"
318266..318463
/gene="ECs1408"
318266..318463
/gene="ECs1408"
/translation="unknown,similar to hypothetical proteins e.g. L0009
[Escherichia coli] gi|3414877|gb|AAC31488.1| percent
identity 87 in 65 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34831.1"
/db_xref="GI:13360868"
/translation="MKSLTTFETALDILIAWLQDNIDCGSIIIFDNDKDTDSALLPC
IEQAREVTRLRHQLLHQNR"
318548..319390
/gene="ECs1409"
318548..319390
/gene="ECs1409"
/translation="unknown,similar to hypothetical proteins e.g. L0010
[Escherichia coli] gi|3414878|gb|AAC31489.1| percent
identity 81 in 111 aa, ydia [Plasmid Colib-p9]
gi|4512489|dbj|BAAY5138.1| percent identity 37 in 265 aa,
L0012 [Escherichia coli] gi|3414880|gb|A"
/codon_start=1

```

```

/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34832.1"
/db_xref="GI:13360869"
/translation="MSDITSRPEVANGHTDVICSTIRHILAYRKSTLLDITLIRQ
LAESVLTESIGGTAPDAMKQDFRCGWLMEKPEMATAITRNLRDREIWRDLWQRS
LAEISLMDTQARDTWYRLESDNFPEISEANLITFEQLHQNKDFVFERGVINVRGLN
WNTKNCPCFGSKIIIVNNLVRWDWGFHLITGQOADRDLADLERMLHLFSGKPIPDNR
GNITLRDDHITQSVQKSEVEDEMFISIRYFKKSAHITFRKPELVDRLNDIIAKHYD
MLAV"
complement(319540..319627)
/gene="serX"
/translation="anticonodon: GGA, Cove score 68.6"
/product="tRNA-Ser"
/complement(319540..319627)
/gene="serX"
319822..320799
/gene="ECs1410"
319822..320799
/gene="ECs1410"
/translation="similar to YCDW_ECOLI gi|1787270 percent identity
98 in 325 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative dehydrogenase"
/protein_id="BAB34833.1"
/db_xref="GI:13360870"
/translation="MRINAFQRFSGVSMDFIYHPTTDTQWIEALKAIPQARVRAW
KSGNDSADYALAWPPVPEMLAGRLKAVFALGAGVDVSIQKQAHPEMLKPSVPLFR
LEDTGMEQOEYAVSQVLHWRFRDDYRICOQNSHWQPLPEYHREDFITGILGAGVL
GSKVAQSLQTRFPLKCSWRSRKSPQFAGWEELSFAFSQCRVRLNLLPNTPETV
GIINQOELKLPDGCAYLLNLARGVHVVEDDLAALDSGKVGKAMLDVFNREPLPESP
LWQHPRTVITPHVAITRPAEAVEYISRTAQLKGERVCGQVDRAGY"
320854..321591
/gene="ECs1411"
320854..321591
/gene="ECs1411"
/translation="similar to YCDX_ECOLI gi|1787271 percent identity
99 in 245 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34834.1"
/db_xref="GI:13360871"
/translation="MYPVDLHMTVASTHAYSTLSYIAQAKOKGKIKLFAITHGPDPM
EDAPHWHFIMNRIMPRVVDGVGILLRGIEANIKNVGDEICSGKMFDSLIIITAGFHE
PVFAPHDKATNTQAMTATIASGNVHIISHPGNPKYEIDKAVAAAKHQAVALIEINNS
SFLHSRKGSDNCRVAAAARDAGGWALGSDSHTAFTMGFEFEBECLKILDAVDPPPER
ILNVSPRLNLFLESRCGMPIAEFADL"
321615..322169
/gene="ECs1412"
321615..322169
/gene="ECs1412"
/translation="similar to YCDY_ECOLI gi|1787272 percent identity
99 in 184 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative oxidoreductase component"
/protein_id="BAB34835.1"
/db_xref="GI:13360872"
/translation="MNEFSILCVLSLYRQDPDLLVPLFTLIREGKLAANWPLEQ
DELLTRQKSCMDTQSYADYNALFIDGECAPYPSAWDEATEAEARFLSERGMPL
RTMAPLTRDAISAMWDELEDSEE"
322223..322762
/gene="ECs1413"
322223..322762
/gene="ECs1413"
/translation="similar to YCD2_ECOLI gi|1787273 percent identity
99 in 184 aa (Conserved in E.coli K-12)"

```

```

98 in 179 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34836.1"
/db_xref="GI:13360873"
/translation="MCSISFMAAFSAIMRGMNILLSIAITTGILSGIMGWAVSLGLL
SWAGFLGCTAYFACPOGGLGLAISAATLLSGVVMWVIYGSALAPHLEILSYVITG
IVAFIMCTQAKOLLISFYVPCFTIGACATFAGOGDKLVLPSLALGLIFGYAMKNSGLW
LAARSAKALREQEIKNKA"
complement(322826..323659)
/gene="ECs1414"
CDS
complement(322826..323659)
/gene="ECs1414"
/notes="similar to CSGG_ECOLI gill178727A percent identity
100 in 277 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curli production assembly/transport component
CsgG"
/protein_id="BAB34837.1"
/db_xref="GI:13360874"
/translation="MORFLFLLVAVMLLSCGLTAPPKEAARPTIMPRAOSYKDLTHLPA
PTGKIFVSYNIQDEGTQFKYPASNFTAVPQSATAMLVTAALKDSRWFIPLEROGLO
NLLNKKIRIAOENGTVAINNRIPLOSLTAANIMVEGSIIGYESNKVSGVGARYFG
IGADTYQOLDJAVNLRVNVVSTGELSSVNTSKTILSVEVQAGYFRFDIYORLLEGE
VGYTNEPVMCLMSAIEGTGVIPLNDGIDRLQDKAERQNDILVIRHMSVPPE
S"
complement(323686..324102)
/gene="ECs1415"
CDS
complement(323686..324102)
/gene="ECs1415"
/notes="similar to CSGF_ECOLI gill178727S percent identity
100 in 138 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curli production assembly/transport component
CsgF"
/protein_id="BAB34838.1"
/db_xref="GI:13360875"
/translation="MRVKHAVVLLMLISPLSWAGTMTQFRNPNFGNNGAFLINS
AQAQSKDPYSYNDDFEIETPSALDNFTQAIQSILGGLLSNINTGKPGRMVNDIV
DIANRGQLQNAVTDRTQTSTIOVSGIQQNSTDE"
complement(324127..324516)
/gene="ECs1416"
CDS
complement(324127..324516)
/gene="ECs1416"
/notes="similar to CSGE_ECOLI gill1787276 percent identity
100 in 129 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curli production assembly/transport component
CsgE"
/protein_id="BAB34839.1"
/db_xref="GI:13360876"
/translation="MKRYLRWIAAEFLPAAGNLHAVEVEVPGLLTDHTVSSIGHDFY
RAFSKWDSDYTGNTLINERPSARWGSWITITVNDQVIFQTLFPLKRFERTVTFAL
IOTEALNRQINQALLSTGDLAHDHF"
complement(324521..325171)
/gene="ECs1417"
CDS
complement(324521..325171)
/gene="ECs1417"
/notes="similar to CSGD_ECOLI gill1787277 percent identity
99 in 216 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="transcriptional regulator for 2nd curli operon"
/protein_id="BAB34840.1"

```

```

/db_xref="GI:13360877"
/translation="MFNEVHSIHGHTLLLTTPKSLQATALLQHLKOSLAITGKLHNHIO
RSLDDISSGSIILLDMMEADKKLIHYVQDTLSPKNNNKKILLNTPEDYPYRDIEWNP
HINGVFYAMEDQERYVNLQGVLRGEYETQKIASYLTHSGNRYRNSTESALLTHRE
KEILNKLRIAGSNNEIARSLFISENTVKTHLNLFKIAVKNRKTQAVSWANDLRR"
complement(325583..325822)
/gene="ECs1418"
CDS
complement(325583..325822)
/gene="ECs1418"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34841.1"
/db_xref="GI:13360878"
/translation="MSAEFLTHPRVVFVFFCYSQSIKTHVNTLLSTOKIFCFLNLN
RFLAKNTRQSFPLPVADCCDMSCAQTNLVFLQ"
325925..326380
/gene="ECs1419"
CDS
325925..326380
/gene="ECs1419"
/notes="similar to CSGB_ECOLI gill1787278 percent identity
100 in 151 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="minor curlin subunit precursor CsgB"
/protein_id="BAB34842.1"
/db_xref="GI:13360879"
/translation="MKNKLLFMMLTILGAPGIAAAAGYDLANSEYNFAYNELSKSFN
QAATIQOAGTNNNSAOLROGGSKLLAVVAQSGSNRAKIDQTDGYNLAYIDQAGSANDA
SISQGAYGNTAMTIQKSGGNKANITQYGTQKTAIVVQROSQMAIRVTQR"
326421..326879
/gene="ECs1420"
CDS
326421..326879
/gene="ECs1420"
/notes="similar to CSGA_ECOLI gill1787279 percent identity
96 in 152 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curlin major subunit CsgA"
/protein_id="BAB34843.1"
/db_xref="GI:13360880"
/translation="MKLLKVAATAIVFSSALAGVVPYOGGGGNGHGGGNNSGPNS
ELNIYQGGNSALALQADARNSDLITOHGGGNGADVGQGGDDSIDLTQRFGNSA
TLDQWNGKDSHMTVKQFGGGNGAAVDQTASNSTVNVTVQVGFGNNAHQY"
326938..327270
/gene="ECs1421"
CDS
326938..327270
/gene="ECs1421"
/notes="similar to CSGC_ECOLI gill1787280 percent identity
98 in 110 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative curli production protein CsgC"
/protein_id="BAB34844.1"
/db_xref="GI:13360881"
/translation="MNALLLLAALSSQITFNTTQOGDMYTIPEVTLTQSLCRVQIL
SUREGSSGSGSQTKQEKTLSPANQPIALTKLSLINSIPDDRKIVTVVSDGQSLHLSSQ
WPPSSEKS"
327391..327702
/gene="ECs1422"
CDS
327391..327702
/gene="ECs1422"
/notes="similar to B1044_ECOLI gill1787281 percent identity
98 in 103 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"

```

```
/protein_id="BAB34845.1"
/db_xref="GI:13360882"
/translation="MRPFLNSLMGLSPFFTAIAAGSTAGGVHIFYGQIVEPACDV
STOSSPVEMNCPONGSVPGKTYSSKALMSGNVRNAQIASVKVYLDKQKKLAVNNIEY
N"
BASE COUNT      83181 a 76892 c 84754 g 82946 t
ORIGIN
AP002554 Length: 327773 February 15, 2002 13:42 Type: N Check: 6396 ..
Found using 'seq4' (pappu403.key)
...
9033  ATTTGCCCTTCTTACCGCTTTCACCACTTCACTGTGAAGACTTAATAGCGGAACGTTAG
          9083 9090
          |-----|
9093  CTGTGGGATAATGACACCTTGTTTCCCGGTTAACTCACGTTGCTGCG
          |-----|
          14396 14403
14346  GTATGTTAAGGCACAGCAACCGAAACGGGCTTCGCGAATTTGCGGGTAACGTTCCGC
          |-----|
          14396 14403
14406  TCGCTGGACATATTGATACCAACCGCAGCCAGCTCAGCATCACGGTTG
          |-----|
          69831 69838
69781  GCGGCTCTGAACATCACTCAACGCTGAGCGGGGTATCTTCATGCTCGACGAACGTTCC
          |-----|
          69831 69838
69841  TGGCACCGGATGAAGTGAACGGGATTAAACGTGAATAACCGTTATGAGTT
          |-----|
          122012 122019
121962  CTCAACGCCCAAAATCCCGGCACTGAATGACACAGCGTAAACACAGATCCGAACGTTCT
          |-----|
          122012 122019
122022  GCGCAAAACCGGAGTGATTAACCCGCCAGCTTGACGGGCATGGGTTCATCA
          |-----|
          177928 177935
177878  CTACAAGAGGGGAGAGCGCATGACGCAAGCGCGAGTGAACCAACGGGTAAACGTTTCGCG
          |-----|
          177928 177935
177938  CGCAGTAAGCGCGAAGAAAAAAGCGATTCTTAGCGCAGCACTGGACAC
          |-----|
          195915 195922
195865  TTCATAGCTTCCGCATAAACAATGGCTTGCCCTAAATGCCGGAACAATATCGAAGCTTTT
          |-----|
          195915 195922
195925  TATAGGGTTGTAGTATGGGGTATCGTGTCTGTATTGTTTCGTAATA
          |-----|
          207278 207285
207228  ACTGGAGTTATAGACTTTTGCACCTTCTGCGCCTTTAGCTTCGACCGGCGAAGCTTC
          |-----|
          207278 207285
207288  ACCCAGATAAGTCCGAAGTAATGCCGCCACCATGCACAGCCATCGC
          |-----|
          219715 219722
219665  CAGGCCAGGAGTATCTCTCTTCTCTGACCGACATTTGCGATATCTGCGCGCAACGTTT
```

```
219725  GAGGAAGAGCTGAAGTCGTTACGTGATATTCACGCACGCATTCACCGC
...
224359  ACGAAGCCTTGGACAGCAATGCTGGATAGCATTGCCTGGCGGGCGCAACGTTCCGC
          |-----|
          224409 224416
224419  TTTATCTGACCGGTAATTGGGCCGCGTATCGACGGGCTGGATTACAAAT
...
319834  TTGCGCACACAGCCTTTTCGGGAGTCAGTATGATATCATCTTTTATCAACCCCAACGTTCCAT
          |-----|
          319884 319891
319894  ACCCAATGGTGGTAGAGGCACTGCGCAAGCTATTCTCTCAGGCAAGA
...
-----
6 matches found in sequence:
ap003361 : TOIG of: ap003361 check: 3057 from: 1 to: 347235
(from "immun_ge.seq")
TOIG of: ap003361 check: 3057 from: 1 to: 347235
-----
LOCUS      AP003361 347235 bp DNA BCT 30-MAY-2001
DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 4/9.
ACCESSION AP003361 BA000017
VERSION   AP003361.2 GI:14246761
KEYWORDS
SOURCE    Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM  Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE 1 (sites)
AUTHORS   Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
Mizutani-Ui,Y., Kobayashi,N., Tanaka,T., Sawano,T., Inoue,R.,
Kaito,C., Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S.,
Yabuzaki,J., Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K.,
Yoshino,C., Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and
Hiramatsu,K.
TITLE     Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL   Lancet 357 (9264), 1225-1240 (2001)
MEDLINE   21311952
REFERENCE 2 (bases 1 to 347235)
AUTHORS   Ohta,T.
TITLE     Direct Submission
JOURNAL   Submitted (28-PEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohta@sakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
COMMENT   On May 29, 2001 this sequence version replaced gi:13875305.
FEATURES
            Location/Qualifiers
            1..347235
                /organism="Staphylococcus aureus subsp. aureus Mu50"
                /strain="Mu50"
                /sub_species="aureus Mu50"
                /db_xref="taxon:158878"
            35..1021
                /gene="SAV0992"
            35..1021
                /gene="SAV0992"
                /codon_start=1
                /transl_table=11
                /product="oligopeptide ABC transporter ATP-binding protein
homologue"
```



```

/product="GTP pyrophosphokinase"
/protein_id="BAB57163.1"
/db_xref="GI:14246771"
/translation="MKQLSYKVSKEIKYIYTCYNEIIPKRDIITSMICDCIQNAG
IDLEVFDLQKSLTESKIDLIHAREMETEQAPSLVFFSEDVHEEGLKVEGLYPVH
IYTIINELMGKPLEKKNLPKLETYIQOQQLVTMEELTIYEWPEKLLNKLKLAIQ
complement(10575..10901)
/gene="SAV1002"
complement(10575..10901)
/gene="SAV1002"
/codon_start=1
/codon_table=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57164.1"
/db_xref="GI:14246772"
/translation="MNMAGELRMENKSRREDINLSPVSKIEIYFFDPFSSDCFKLSA
ILSKRIEYQYIRIRHILNPSLKVLTCKQAQSTSFNDNALAYKAAELQGRVEDRKO
LYTHML"
complement(10918..11283)
/gene="SAV1003"
complement(10918..11283)
/gene="SAV1003"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57165.1"
/db_xref="GI:14246773"
/translation="MTTPYDIIGKEALYDMIDYFTLVKEKDERLNLHPGDFAEFSR
KQQLFTQFLGGPNITYEEHGPMLRKRHMDFITTEFERDAWLENMQTAIRNAFFPQG
VGDLFERLRLTANHVMNS"
complement(11387..11980)
/gene="SAV1004"
complement(11387..11980)
/gene="SAV1004"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57166.1"
/db_xref="GI:14246774"
/translation="MATNHEIEFKOMTASIYNKLOEKYFKDSVLFKQVNYIIDPDF
KLEHRSAIRIRVKDQMYETLTPAKVGLLENYIVDIKPEMNLIIISNDNLDDIRQ
IIVQFGVKDQTLTILGALTYTQETKYKGDLLVLQKSEYFDADYELEFEVFKDYNGQ
LOKQSLNELNELHQLPNKQVRFKKETLSNNIN"
12166..12513
/gene="SAV1005"
12166..12513
/gene="SAV1005"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57167.1"
/db_xref="GI:14246775"
/translation="MRLYINEIKDKDILCYTEDSIKGLSEVGOMLVSDNYAFAYT
LDDGAYAYLIFVQETWTMLHENTTKKIIINDELETFEQELTYILDNIKGNNNYK
EFVATVEETFDIE"
12530..13165
/gene="SAV1006"
12530..13165
/gene="SAV1006"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="BAB57168.1"
/db_xref="GI:14246776"
/translation="MNQWQFLTPYQAVDELKVKLGMRKQVEVGEOASPEFTVTR
VKPIASIDKANKROIIPFDRLREMYDIAGLRMQCFVEDIDVYVNLIRQRDKFVIE
ERDYIRNTKESGVRSYVHIIIEYPIETLQCKFILAIEIQRILAMPWATIEHTLRKY
DGAYPEIGHRLERARAAAYLLDEEMSEIKNEIQEAQKIYTKRRSKKHEND"
13182..13991
/gene="SAV1007"
13182..13991
/gene="SAV1007"

```

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57169.1"
/db_xref="GI:14246777"
/translation="MRYTILTKGDSKSNALKHKMMYMKDFRMIEDSENPEIIVSVGG
DGIILQAFHQISMLSKVAFVGVHTGLGFTADMLPHEVEKLLIIEINSEFQVIEYPL
LEIMRYNDNGYETRYLALNEATMKTGSTLVVDVNLRGKHFRFRGDLGCVSTPSCG
RMTYKALGALIHPSLEAMQITEIASINNRVFRVTGSLVLPKPHHTCLTISPVNHDTI
STYIDHVSIKHKNVNSIQYRVANEKVFARFPPFPKRVHDSPTSSDEER"
13988..14842
/gene="SAV1008"
13988..14842
/gene="SAV1008"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57170.1"
/db_xref="GI:14246778"
/translation="MKPKYHISQOETVKAFLARHDFSKTVSIAKNNGALIVNDEPYT
VRKQLPNDILEIHLREIPSNLIPYARKLEVLYEDAFIIVTKPNNQCTPSREHP
HESLIEQVLYHCOEHGENINPHIVTRLDRTGTGIVIFAKYGHITHLPSKVNLLKIYTC
LVYKTHTSGLIEANIRRSKDSIITREVASDGKYAKTSYEVINQNDKYSLCKVHLHFG
RTHQIRVHFQHIGHPIVGDSDLYGDFHDKINGQVLOCTQIYIFVHPINKNNIYIIDYKQ
LLKLFNQL"
14863..16248
/gene="SAV1009"
14863..16248
/gene="SAV1009"
/notes="similar to Mg2+ transporter"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57171.1"
/db_xref="GI:14246779"
/translation="MSMNTDEKERVQOEELYDETLLOYLENDIDIDQFRODFELAHYTE
QSEYFEDTDENRQKIFQLSPREEVANFFDOLDIDDEYELLFDDKMNATFASHLEEM
SYDNAVDILLNLTPKPVASLLTLMNKDDANEIKALLHYDEDTAGGIMTTLSLKHAHT
PVKEALLVKAQAPDAETIYVIEVDDDGKLVGLSDRLIVAEINDAYIEDIMNERVI
SVNVAQDEQVADQMRDYDFMAVPVIDYQEHLLGIITIDDLDMVDEASEDYSLRAG
VSDIDTNDISIITALKRLPWLIILTFGLMTATILGRFEKTLLENVALLAFAPIIISG
MGSNSTQSLAVSVRNITTEINEQSKFRALREAGSGVLSGVVCSITLTIIVAIYH
QPLALLIAGSLTCAMTVGTFGVSGMPLLMKNLUNIDPAVASGPFITINDIISMLIYF
GLATSFMAVLI"
16258..18102
/gene="SAV1010"
16258..18102
/gene="SAV1010"
/codon_start=1
/transl_table=11
/product="Na+/H+ antiporter homologue"
/protein_id="BAB57172.1"
/db_xref="GI:14246780"
/translation="MEELSIVIVVLAFLTPIIVNRLNINFLPVVAETILMGIVIGNS
FLNIVERDSILNLTSGFIFLMFLSGLEIDFKAKDKRARGQNDDESSIPGHLNL
ALTUFAFIMIIISLLAYFKVLGDDVLLMIIITISLGVVVPVTKEMNIMRTTIG
QETLLVAVLADLVMTLITVYGAINGQSGTWILGILVFTVTAISYILGVQFKRMSFL
QKLMGDTTOIGTRAVFALIILLVLAEGVGAENILGAFLGAVVSVLNDPEENVEKLD
SPCYGFFPIFFIMVGDNLNPSLKEPKLIITIPILIVAFIISKLIPVMFIRRFDM
KTTIASAFLLTSTLSLVIAAKISERLNAISAEISGILSILSAVITCVFPIIFKFLFP
VDFENRKEVSLIGRQNLTIPIAQLTISQLDVTLYIRKDLSDRRQLSDSDITMIEIA
DYEQDVLRLGDFDRDIVVVCATNDDDINRKLAKAKAHOQVIRVYKLESTDDTDELVD
SGIEIFSSYLSNKLILKLTIPNMLNSVETSLYEIQMLNYKENTQLRNFPEFGG
DIIFVRIIRNNEISVPHGDTQLRYGRDLITVGAKEYVDELKQLEFYF"
18380..19150
/gene="fabi"
18380..19150
/gene="fabi"
/notes="SAV1011"
/codon_start=1
/transl_table=11
/product="trans-2-enoyl-ACP reductase"
/protein_id="BAB57173.1"

```

```
/db_xref="GI:14246781"
/translation="MLNLNKTYVIMGIANKRSIAFGVAKVLDQLGAKLVFTYRKERS
RKELEKLQLOPQEAHLQIDVQSEEVINGEQIGKDVGNIDGVYH5IAFANMEDL
RGFSETSREGFLAODISSYSLTIVAHEAKMPEGSIIVATYILGGFEAVQNYVM
GVAKASLEANKVYALDGLPDNIRVNAISAGPIRTLSAKGVGGFNTILKEIERAPLK
RWDQVEGKTRAYLLSDLSGSGVTGNIHVDSGFHAIK"
complement(19347..20432)
/gene="SAV1012"
complement(19347..20432)
/gene="SAV1012"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57174.1"
/db_xref="GI:14246782"
/translation="MLNKVFRTGIALIMFLIKLFMEVHEVFTPIATIGSVLPFP
LISGFLFYICLPONLEKVGPRWASITIMLAFIIGLIVAFVAPILISINFLI
SQTPALQKAEQIKFALAQOMKLPEDVTSRITNVKSMGDBGVNLISLSQIITSLI
STIFLIMYFFELIYMLKDHEKPIPAVAKFFGERKVFFVDLLDNLFTLKSYIQGV
TVSVILGYIYIGSIDLPYILPLVAFAGVANLIPFLGSLWLSFAPAAIIGIIDSPTT
FWVCITILIAHOLEGNIITPNVMGKSLSTHPLTITIIWILAAGDLGGFTLVLIAPVLY
AVLKTVSNLEKYRQRIIDKANSNVKD"
20837..22342
/gene="SAV1013"
20837..22342
/gene="SAV1013"
/notes="similar to Na+/H+-dependent alanine carrier
protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57175.1"
/db_xref="GI:14246783"
/translation="MLITGVLEFLMRFPQVRHFKEMIRLMFOGEKSPNGISSFQAIA
MSLAGRVGNIGVSTAIFIGGPAVFWMTAFLGASAFTESTLGIQIFKRVENNE
YRGPAIYIEYGIGGKFGKIYIFAIVTIIISGLLPGVQSNAIASSMHNATHVPQW
LGAIVVYILGHTIFGGVRSIANVAVPFWMAIYILMAVIIICINIOEVPALFALI
FKSAFQLQSAFGIIVGAMIEIGVKRGLYVNEAGQGTPHAAAAAESHPSKQGLVOAF
SVYIDFLVCTATALIILISGTYNTDGTNVANGTPHLIKDGGIYVENATGKDYSGTA
MYAAGIDKAFHCGSYQDPDTPSGVGSYFIAPALFFAFTTILSYIYIETNVAYLTR
QNQNVSSIFINIRVYIILFATFYGAVKTADVAMFGDLGVGLMWNLIIAIWLHKP
AVNALKDIEIQKRLGNGYNVYQDPNKLPNNAVFLKTYPERLKOARAKK"
22487..23245
/gene="SAV1014"
22487..23245
/gene="SAV1014"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57176.1"
/db_xref="GI:14246784"
/translation="MSEFTGKINKVILNSILNRDVTISIYLPESYNQLVKYNVILC
FGLDLFLRGRIQRTYESLKEARIDAIIVGHYEDVDKRREEPHPQSRSHLTIOS
VKKEILPFDSTFSTKVGNRALLVDGSLAGSALTALTPTPTIFSRVAMLSPHSDDK
VLDKNOCANKEQLTIIWHVIGLDEKDTIPTNGKRADELTPNRELAEQIKKYNITYY
DEPEGCHONKWKPLLSILLYFLRNTDDQHYE"
23439..23948
/gene="SAV1015"
23439..23948
/gene="SAV1015"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57177.1"
/db_xref="GI:14246785"
/translation="MILGLALIPSKSQEAVDSYRKRKYQKYSRIKPHVITKAPFIEE
DGLDLSVIEQVRARINGIPAVEVHATKASSFKPTNNVIYFKVAKTDDLELPFRNGE
DFYGEAEHVFPHTTIAQGLSQEFEFIDFGQVALAGVDHKEIIDELTLLRFDDDEDKW
KVLETFKLA"
complement(24061..25251)
/gene="SAV1016"
complement(24061..25251)
/gene="SAV1016"
```

```
/notes="similar to multidrug resistance protein-related
protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57178.1"
/db_xref="GI:14246786"
/translation="MQDSSLNNYANHKNFIMLIILFLMEFARGMYILSYINFLPVTY
SIAVAITSIAFSIHFTADASTNEVIGFLKFGTKIVLTGTFILAFSLFLVTFWFPAS
PFIYIFSAMLGIATVP IAWIMLSSVEEDKRGQMVYFESWLLGLLVGMVFNLLIK
VHETRAFMSIYVLLIAWILLYEVDVKLTNYNTRPVKAOLRQIVDVTKRLLLPGLIL
LQGAARALPILPTATKVINVSTIETVVAIIIGIGGCASVLMFLSKLIDNRSRPM
YGVILSGFILYMILIFTSMIWNHIIAIGLMYIGLIPAWNTFMARFKISDQQ
EETWGVFNSIQGFGSGMIGPLFGLGITQFTNDLNTNTFFSALIFLVAVFYGSYFIANR
EKAK"
complement(25229..26404)
/gene="SAV1017"
complement(25229..26404)
/gene="SAV1017"
/notes="similar to cell wall synthesis protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57179.1"
/db_xref="GI:14246787"
/translation="MVTQNKKILITGSGFGNHMQVTQSIYNQINDMNLDSLVIEDH
LFMEAHPILTSICKKWYINSFKYFRNNYKGYFSRPDKLDCFKYKYGLNKLNLILIK
EKPDILLTFPTVMSVLTQFQNTIPVATVMTDYRLHKNIPTPSYRYVATKETQK
EFDVIGIDPSTVKVGTGIPDNKEETPINQOWLIDNNDPKQOTILMSAGAFVSGKGF
DTMITDLAKSANAQVVMICGSKELRSIAKFSNENVLILGTYTKHMENWASSOL
MITPGGIIITEGFARCIPIFLNPAGQLENALYFEERKFGKIADTPPEAKIVAS
LTNGNEQLTNMISTMEQDKIKYATQTICRDLDDLIGHSSQPEIYGVKPLYARFFVK"
26837..28321
/gene="murE"
26837..28321
/gene="murE"
/notes="murE"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylmuramoylalanyl-D-glutamate-2,
6-diaminopimelate ligase"
/protein_id="BAB57180.1"
/db_xref="GI:14246788"
/translation="MDASTLKKVKVKVGLGSEQQIDDIITDSRTAREGSIFVASVG
YTVDSHKFCONVADOGCKLVVNKEQSLPANVTQVVPDTRLVASILAHPLYDYPSHQ
LVTFVGTGCKTSIATMIHLIORKLOKSAYLCTNGCFQINETKTCGANTTPETVSLT
KKIREAVDGAESMTLEVSGLVLRGRVDFVAIFSNLTQDHLDFHGTMEAYGHA
KSLLSQLGDLSDSKERYVILNDDSFSEYLRVTPTVEYFVSIGIDEEAQFMKNQIESL
QGVSFDFPTPGTYPVKPSYVGFKNISINMAAMIAVMSKGTSLETIIKAVENLEPVEG
RLEVLPDPLDILIDYAHADGMNKLIDAVOPFVKOKLIFLVGMAGERDLTKTPMG
RVACRADYVIFTPDNPANDPKMLTAELAGATHONYIEFDDRAEGIKHAIDIAEPD
TVVLASKGREPYQIMPGHIKVPHRDDDLGLEAAKKFGGGFPVQ"
28311..28562
/gene="SAV1019"
28311..28562
/gene="SAV1019"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57181.1"
/db_xref="GI:14246789"
/translation="MINKRFIDEGKTIIDVYLFELANNQIIIAIPDWFSYQMAMTLDE
ETCFEAILMQLFVFKEEEEAEASLTQDWIETYYKREK"
28604..28780
/gene="SAV1020"
28604..28780
/gene="SAV1020"
/codon_start=1
/transl_table=11
/product="peptide chain release factor 3"
/protein_id="BAB57182.1"
/db_xref="GI:14246790"
/translation="MISHPDAGKTTLTLEKLYFSGAIREAGTVKGGKTKFGDQVTGN
```

gene

ESLNQRGVGYFLYT"

CDS

28952..30169
/gene="SAV1021"
28952..30169
/gene="SAV1021"
/codon_start=1
/transl_table=11
/product="peptide chain release factor 3"
/protein_id="BAB57183.1"
/db_xref="GI:14246791"

gene
CDS

35328..36842
/gene="SAV1025"
35328..36842
/gene="SAV1025"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57187.1"
/db_xref="GI:14246795"

gene
CDS

35328..36842
/gene="SAV1025"
35328..36842
/gene="SAV1025"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57187.1"
/db_xref="GI:14246795"

gene
CDS

35328..36842
/gene="SAV1025"
35328..36842
/gene="SAV1025"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57187.1"
/db_xref="GI:14246795"

gene
CDS

35328..36842
/gene="SAV1025"
35328..36842
/gene="SAV1025"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57187.1"
/db_xref="GI:14246795"

CDS 39591..39767
/gene="SAV1029"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57191.1"
/db_xref="GI:14246799"
/translation="MTVAEYGNIVEFMDGLGRVEXKINDNSVIVDLTIMENFNDLDP
EKTVINHKRYKIVE"
39782..40384
/gene="SAV1030"
39782..40384
/gene="SAV1030"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57192.1"
/db_xref="GI:14246800"
/translation="MNKISKALTWFTIIPHLILFMGGEHQEYWYLYTGIMLIAG
ISYVFQDIESKRLTISGVGIITAIILIMLQLLFSLTISLSYSSLIKELARTGVN
WKQMVLVTLLFVIPCHELYMRTVLOKELTHFSIPKWIAILITAICSSSLFIYLDNWWI
VTFIFVAQVILSLSYEYTRRIATTSVAQIVAILLLIFNA"
40593..40754
/gene="SAV1031"
40593..40754
/gene="SAV1031"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57193.1"
/db_xref="GI:14246801"
/translation="MSKGGWLWHYLTLLKRAYCDIYCKRITFNVKFYVYRHPHWY
SSSNKFKQ"
41411..41683
/gene="SAV1032"
41411..41683
/gene="SAV1032"
/note="similar to lactococcin 972"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57194.1"
/db_xref="GI:14246802"
/translation="MFGTILYTLALGLSTAAYASTEYAEGGTWSHGVSQYVWSYYY
HGKHGGAIGAIGYRSFGYTRAGVAKASATKHCWVNRVYNYI"
41727..43691
/gene="SAV1033"
41727..43691
/gene="SAV1033"
/note="similar to lactococcin 972 immunity factor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57195.1"
/db_xref="GI:14246803"
/translation="MKWEKLIDVTFTLIALILFVYTYKENEELPDTKYPIATVDM
NKKYSKNEYIKRINQFANENVAIYKSTSNYNKNVDKIYVFNKSKATTIPFNAYK
NTHYLSDDLELLKKDIKGSYFVKDKNFDVSKFNEFLKGYGVTAEYSKIDHMTIAGVIK
QMNIEVPSALLIVYIYIYFEKNINFKAYAKYLNGLFTLRKLIIFENFSKCTYVWTL
IISQILLTSLVILNYGNLDFILRVLLSCLFTILTSVNLMTFLMLNLNIAHM
ISGQHKFTIRFINTVCKSLILLVIASVMIENTSVIKDLNKIKETEKYNVLDYITI
EPAPYHETQSLIDNMLRSEQLKYVSEANNAILFKPGSDVNDNFSDEGNVILVN
NQFWSIYKQFQDPIRKQKNNEVVIPOKPHAMRNEINQAYHSWFEPVQNNKEN
KLSIQFINKNDRIFSEDARDSRHLSEFTEAPIIVNVQASDLSNDFYAMISOGGYLFK
NYDALVKNIENYHLDGEISGITYKDSVMYEMHNNKLTVLNFSQIIITAILIILIL
FDVKYFQGRKLLVKKLIGYSTLRANYQYLLINNIVVVFIGILTNVILHSQYIMMI
FATILVQLLQICLSYHGRFNEVKEF"
43694..44014
/gene="SAV1034"
43694..44014
/gene="SAV1034"
/codon_start=1
/transl_table=11

gene
/product="conserved hypothetical protein"
/protein_id="BAB57196.1"
/db_xref="GI:14246804"
/translation="MKFIIAILLGLIVSITIAFTIIHHPILDRENPFLKTEYSYAKVP
KGTQYVNIITAYISERGEKLDYKLTFFNGFSPSRTYVEIKHKGQYVISITYVEKEDIPKE
VROE"
44011..44652
/gene="SAV1035"
44011..44652
/gene="SAV1035"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57197.1"
/db_xref="GI:14246805"
/translation="MIELKDLTIQKGNIIHLKLNLFQCGKSYALLIGKSGCGSKSTLL
NTIAGLEKTKQYVFNQLEQKSNFYRKDGLYLFQNYGLINDLITVNEMLDIGLAYK
KTSKKEQKQIRYIEQFGLSNLSRKRVHTLSGGEQORVALIRMLKDKPTVLADEPT
GALDPKRTGMIIOQLFDLVNENKVLILATHDMAIANQCDEIIDLEQYSKVASM"
complement(44740..44985)
/gene="SAV1036"
complement(44740..44985)
/gene="SAV1036"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57198.1"
/db_xref="GI:14246806"
/translation="MRTYNIIGILSCLISFIIMALPMIWTASALWFFPGAIMILL
SIVIVFCYIKTKNQLHLLIVLNIILFFSLPLLS"
complement(45397..45753)
/gene="SAV1037"
complement(45397..45753)
/gene="SAV1037"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57199.1"
/db_xref="GI:14246807"
/translation="MKLGMILLIRWMLATSFVLHGTILKFVDLDCGTIHFSLSLGPPTIA
VLMISGEVVGGLALIGFLTPYVNVFYICVMLGSIFTKLTRGYVSGYEFEIHIHVMN
46241..47200
/gene="SAV1038"
46241..47200
/gene="SAV1038"
/note="similar to ferrichrome ABC transporter"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57200.1"
/db_xref="GI:14246808"
/translation="MNRNIVKLVFVILVAVAGCGQDKTEKTEMTTIKDELGTGEK
IKNPKRVVVEYSFADYLAALDMKPVGIADGSTKNITKSVRDKIGAYESVGSRPQP
NNEVISKLKPLDIADISRRHKIKSELSKIAPTIMLVSGTDYDNANIEAFKTVAKAVG
KKEGSKLEKHDKILAEIRKKIEQSTLSAFAGISRAGMFINNEDTFMGQFLIKMG
IOPEVTKDTHTVHGRGGPYIVLNNEELANINPKVILATDGTCKNRTKFDPAVW
KSLKAKDNKVVDVDRNKLKSRGIITASESMAEDLEKIAEKAK"
complement(47441..47656)
/gene="SAV1039"
complement(47441..47656)
/gene="SAV1039"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57201.1"
/db_xref="GI:14246809"
/translation="MKVNKAIYVIVALFGLGLGIHKFYADKFGQGLHLLFFWTGIPT
IISIIHAIHFVITFKADKDGYYIFPKN"
47821..48372
/gene="SAV1040"
47821..48372
/gene="SAV1040"

gene
/product="conserved hypothetical protein"
/protein_id="BAB57196.1"
/db_xref="GI:14246804"
/translation="MKFIIAILLGLIVSITIAFTIIHHPILDRENPFLKTEYSYAKVP
KGTQYVNIITAYISERGEKLDYKLTFFNGFSPSRTYVEIKHKGQYVISITYVEKEDIPKE
VROE"
44011..44652
/gene="SAV1035"
44011..44652
/gene="SAV1035"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57197.1"
/db_xref="GI:14246805"
/translation="MIELKDLTIQKGNIIHLKLNLFQCGKSYALLIGKSGCGSKSTLL
NTIAGLEKTKQYVFNQLEQKSNFYRKDGLYLFQNYGLINDLITVNEMLDIGLAYK
KTSKKEQKQIRYIEQFGLSNLSRKRVHTLSGGEQORVALIRMLKDKPTVLADEPT
GALDPKRTGMIIOQLFDLVNENKVLILATHDMAIANQCDEIIDLEQYSKVASM"
complement(44740..44985)
/gene="SAV1036"
complement(44740..44985)
/gene="SAV1036"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57198.1"
/db_xref="GI:14246806"
/translation="MRTYNIIGILSCLISFIIMALPMIWTASALWFFPGAIMILL
SIVIVFCYIKTKNQLHLLIVLNIILFFSLPLLS"
complement(45397..45753)
/gene="SAV1037"
complement(45397..45753)
/gene="SAV1037"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57199.1"
/db_xref="GI:14246807"
/translation="MKLGMILLIRWMLATSFVLHGTILKFVDLDCGTIHFSLSLGPPTIA
VLMISGEVVGGLALIGFLTPYVNVFYICVMLGSIFTKLTRGYVSGYEFEIHIHVMN
46241..47200
/gene="SAV1038"
46241..47200
/gene="SAV1038"
/note="similar to ferrichrome ABC transporter"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57200.1"
/db_xref="GI:14246808"
/translation="MNRNIVKLVFVILVAVAGCGQDKTEKTEMTTIKDELGTGEK
IKNPKRVVVEYSFADYLAALDMKPVGIADGSTKNITKSVRDKIGAYESVGSRPQP
NNEVISKLKPLDIADISRRHKIKSELSKIAPTIMLVSGTDYDNANIEAFKTVAKAVG
KKEGSKLEKHDKILAEIRKKIEQSTLSAFAGISRAGMFINNEDTFMGQFLIKMG
IOPEVTKDTHTVHGRGGPYIVLNNEELANINPKVILATDGTCKNRTKFDPAVW
KSLKAKDNKVVDVDRNKLKSRGIITASESMAEDLEKIAEKAK"
complement(47441..47656)
/gene="SAV1039"
complement(47441..47656)
/gene="SAV1039"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57201.1"
/db_xref="GI:14246809"
/translation="MKVNKAIYVIVALFGLGLGIHKFYADKFGQGLHLLFFWTGIPT
IISIIHAIHFVITFKADKDGYYIFPKN"
47821..48372
/gene="SAV1040"
47821..48372
/gene="SAV1040"

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57202.1"
/db_xref="GI:14246810"
/translations="MIRQARPEDRFDAKLVYMWDDMELELYKHLPKDMVLDAIEKS
CVDATRTFQHLIVLVENKQVAGCIISYSGNELSKYKAWELLDLPEEKQVGTPLP
VKAKDDEYVIEFIATFAARGRIATKLLTSLESNTHVKWSLUNCINDINNEAALKLYK
KVGIFSDGQLELYKHYVHLIVK"
complement(48424..49362)
/gene="SAV1041"
CDS
/gene="SAV1041"
complement(48424..49362)
/note="similar to 1,4-dihydroxy-2-naphthodate
octaprenyltransferase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57203.1"
/db_xref="GI:14246811"
/translations="MSNQOQSYSTVKKYWHLMRPHLTLTASVVPVLVGTAASKIYFLGS
EDHKISLFTAMLLACLLIQAAATNMFNEYDYKKGLDDHESVGGIAIVRNGMSEPLY
LRIAIFATILGFLAANSFVLLPVGLVCMVAGYLVYTGPPFTSPTWPGELPSG
VFMGMFIIATFIOTGNIQSYVIMLSVPITVITIGLINMANNIRDRVKDKASGRKTLP
ILGKNASLTFAIMYFIAYAFIVLIIIKPGSLEYLLALLSFPMPVKVIRFRKND
TPTTPMPANAAGKTNITFFGLLYALGIYISALFAGI"
49544..50905
/gene="SAV1042"
CDS
49544..50905
/gene="SAV1042"
/note="similar to menaquinone-specific isochorismate
synthase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57204.1"
/db_xref="GI:14246812"
/translations="MATGVLEDDIVKEIYSSKEWVSVEVKLSQSLDPSTLPHLTONE
ADGRYMLRNDRTSYFGKAIQLKFNKNKQKSIKDWKELKHNTFIHPQSEKHLL
RVVGFQFSHKSDDREFGLNHFLVPLVLTIDNNGTFLTVYKRGSEFTVPAALND
MOLFNNISIDVDEQIGETRNEDIYKDDWROLVVEAIESINNEEKIVLARRLIKFD
KDISPIYILQAYSKEKNSYIFLLESQDSIFFSQTEQLIKYNNKILSTKAVAGTIKR
SQDEDETKNVAFLKDNKNLIEHRFVDSILHDIKAPITELHYDKTPKILKNDHLIY
LYTEIKAPLDDSYISLIDHLHPTPALGGYPKEFAMDFIQEKEFGTRGLYGAPVGIY
IYDDCEFIIVAIRSMILIKAAATLFAAGCGIVKDSDDPSLAETNLKFTPMNALGVDMN
GKS"
50892..52565
/gene="menD"
CDS
50892..52565
/gene="menD"
/note="SAV1043"
/codon_start=1
/transl_table=11
/product="menaquinone biosynthesis protein"
/protein_id="BAB57205.1"
/db_xref="GI:14246813"
/translations="MGNHKAALTQVFTFASELYGVREVVISPGSRSTPLALAPEA
HPNKTWHPDRSAAFFAVGLIKGSRPVAILCTSGTAAANTVPAIESQISRIPLI
VLTSDRPHLSRVGAPOAINOVNMFNYSYBFDMPADDSKETINAIYYQMQLASQY
LYGPHRKHPLNRPDPLTPDLNATELLTSEMKLPHYQKSIDASALRHILNKKGL
IIVGDMQHQEVQDILTYIIDPLADPLSHLRKFPDNPVICTYDLLFRSGDLNVD
FVIRGVPSIKKLQWCKLKKRKEIKYLEQADSEAFVGLIKKTSKDALFISNSMP
DTTINRVSWLEKQWCKLKKRKEIKYLEQADSEAFVGLIKKTSKDALFISNSMP
TRDNLNLIKNIDVYANRANGIDIVSTALGMVHKRITLLIGLDSYHDMNGILLM
SKLNNIOMNVLNNDGGGIFSYLPQKESATDYFERLCTPTGLDPEYTAKLQYDFK
RNSYSSEKNATLLSETSTIYELTINRDNFQHOILYQKLSEMHDTL"
52552..53355
/gene="SAV1044"
CDS
52552..53355
/gene="SAV1044"
/note="similar to prollyl aminopeptidase (EC 3.4.11.5)"
/codon_start=1
/transl_table=11

```

```

/product="hypothetical protein"
/protein_id="BAB57206.1"
/db_xref="GI:14246814"
/translations="MTHYKVEANVTNQVLVLLHGLFSDSRTYYNHIQKTDICHVLI
TIDLPGHCEQSSMDETFWNFDYITLLDRILDKYKDSITLFGYSMGRVALYYAING
HIPISNILESTSPGIRKEEANOGLRRVDDARAKVLDIAGIELFVNDWEKLPFLFOSQL
ELPVEIQHQRILQRLSQSPQKMAKALRDYGTGQPNLWPRLKEIKETPLTIIAGEYDEK
FVQIAKMANLIPINSKCKLISATGHTIHVEDSEDFDTMILGLFLEQND"
53348..54169
/gene="menB"
CDS
53348..54169
/gene="menB"
/note="SAV1045"
/codon_start=1
/transl_table=11
/product="naphthoate synthase"
/protein_id="BAB57207.1"
/db_xref="GI:14246815"
/translations="MTRQWETLREYDEIKYFEGIAKVTINRPEVNAFTPKTVAE
MIDAFSRARDQNVSVILTGEGLAFCSGGDKRGRHGVYVGEDQIPRLNVLDIRL
RIIPKPVIAWKGYAVGGNVLNVGDLTIAADNAIFGOTGPKVGSFDCYGSGLA
RIVCHKKAREIWLICROYNAOEALDMGLVNTVPLDKVEDETVOCKEIKHMSPTALR
FLKAMNADTDGLAGLQOMAGDATLLYYTTDEAREGRDAFREKRDPDFDQPKFP"
complement(54407..54736)
/gene="sspc"
CDS
complement(54407..54736)
/gene="sspc"
/note="SAV1046"
/codon_start=1
/transl_table=11
/product="cysteine protease"
/protein_id="BAB57208.1"
/db_xref="GI:14246816"
/translations="MYOLOFINLVYDTTKTLHLEQTNINLFINFNWSHQLOKSCIRH
GDFSHNQYHILFDIPAHRIKFSIDNEEITYILDYDDTQTHILMQTSSKOGIGTSRP
IVYERLV"
complement(54774..55955)
/gene="sspc"
CDS
complement(54774..55955)
/gene="sspc"
/note="SAV1047"
/codon_start=1
/transl_table=11
/product="cysteine protease precursor"
/protein_id="BAB57209.1"
/db_xref="GI:14246817"
/translations="MNSYSKRVFNIIISIMVSMILSLGAFANNKAKADSHSKOLE
INVSKVPQKVKDLAQOQFAGYAKALDKOSNAKTGYELGEAFKIYKFNGEEDNSY
YPIKDGKIVYITLSPKKNKDDLKSKEDMNSYVKISNFIKADLDQIKDKNSITVLT
DEKGFYFEDGKVLVKATPLPGNVKKEKSAKTVSAKLQELKNTVTPTKVENEAIQ
EDQOYVENTLKNFKIREQQFDNSWCAGFSMAALLNATKNTDTYNADHIMRTLYPEVE
ODLPNGSTFPNOMLEYKKSQGRDIHYQEGYPSVEQDQLTQDNVGMILAQSSQNP
DPHLGHALAVVGNAKINDQEKLIYNWPDWTFELSIQDADSSLLLSLFRNDYNWYGS
MG"
complement(56037..57065)
/gene="sspa"
CDS
complement(56037..57065)
/gene="sspa"
/note="SAV1048
V8 protease, glutamyl endopeptidase"
/codon_start=1
/transl_table=11
/product="serine protease"
/protein_id="BAB57210.1"
/db_xref="GI:14246818"
/translations="MKGKFLKVSFLFVATLTATLVSSPAANALSSKAMDNHFPQOOS
SKQOTPKIKKGNLKLQREHNAVILPNNDRHQITDITNGHYAPVTYIOVEAPTGF
IASGVVVGKDTLLTNKVVVDATHGDPHALKAFPSAINQDNYPNGGFTAEOIITYSGG
DLAIKVFSPNEONKHIGEYVKPATMSSNAFTOVNQTIVGYPCDKPATMWSKCKLI
TYLKGAMQYDLSSTTGCNSPVFNKNEVIGIHWGVNPEFNCAVFINENVRNFLQ
NIEDIHFDNDQPNFDPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN
NSDNPDA"
complement(57593..58747)

```

gene

| | |
|------|---|
| CDS | /gene="SAV1049" complement(57593. .58747) /gene="SAV1049" /codon_start=1 /transl_table=11 /product="Hisc homolog" /protein_id="BAB57211.1" /db_xref="GI:14246819" /translation="MKLSLNSNSKYLRAPSIRFOSFNRMNLLDDCVNLTIGOPDFPMPD VVKKAYIDALNNDKTSYSHNKGLLETRAEASOYEFKNRYHSEYDEELIIVNGASEAID TTLRSIEBGEDEIIPGPIYAGYIPLLEVCGKPIIIDTATQKITPDALESHEID KAVLLNYPNPTGVSATKRNELNIVNKKYIPFIISDEIYAENTSGKHVSFAEFE DIRQDLILGLSKSHSATGIRIGFLGPQYLIDKLTFMHAYNCICANVPAQIACITA LNLEGEAPKYMEAYVERRNYLVSELTKLGEITAOPEGAFYIPFSIKHITDDDFEFC VDLESTHLAIVPGSSFTFGKGFVRISYAYEMDVLEKGMKRLAKLYLNTK" /complement(58943. .59953) /gene="SAV1050" complement(58943. .59953) /gene="SAV1050" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB57212.1" /db_xref="GI:14246820" /translation="MTSLKERDYFDNARAIIILLVVFGHMLQPYTSGDKYLSALYL IYSPHMTFLFISGYFAKNIDKPYILEKISKRLIIVPYMIFFAFFSIYVFLTGKSDLEQ LDPNPFAWFLITLFFHHVILVVRFPYKVLVSIIISIGAGSDNITDLSLSIS RTIVFFPIFYGLFTKKHTAFKNKLLIPVSIITFILFFIIVHPINADWLLGSSP YIISLENGOSIFSFFKRLIYGIILIAMTAFNLNMSTKKLITYIGSKTLIVYLLHGL IIGIVGFEPWYFPDNPISLMTYLDLISIVIIYVLSVTNFVCKWNTNPIINLRSPQK DS" 60106. .60525 /gene="SAV1051" 60106. .60525 /gene="SAV1051" /note="probable ATL autolysin transcription regulator" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAB57213.1" /db_xref="GI:14246821" /translation="MYKOLEKLITLTNNDLNLVNRFGQRTDITSOLELLRLFNVD RLSQYDITMKISREQSIVSRWIKKVLKGYITSQSSSEDLRCKELIITDQARTLISQI NNARCEIARCOCLSEVELDNQLLDKLNQRRISL" complement(60732. .61538) /gene="truncated-atl" complement(60732. .61538) /gene="truncated-atl" /note="SAV1052" /codon_start=1 /transl_table=11 /product="partial autolysin" /protein_id="BAB57214.1" /db_xref="GI:14246822" /translation="MNGQTWYIGKLSNGKLAWIKSTDALAKELIKYINQGTGMTLNOVAQI QAGLQYKQVQRVPGKWTDAFNDVKHAMDTKRLAQDPALKYQPLRDQPONISIDKI NOFLKRGVLVLENOGAAFNKAOMGYINEVYLISSHLETNGTSQLAKGADVNNKVY TNSNTKYHNVFGIAAYDNDPLREGIKYAKOAGWPTVSKAIVGGAKFTGNSVYKAGQNT LYKMRWPAHPGTHOYATDVDWANINAKIKGYIDKIGEVGKYFDIPIQYK" complement(61658. .64489) /gene="atl" complement(61658. .64489) /gene="atl" /note="SAV1053" /codon_start=1 /transl_table=11 /product="autolysin" /protein_id="BAB57215.1" /db_xref="GI:14246823" /translation="MLGVINRMAKKNYKLPWSVALTLVGSAVTAHQVQAETTDQDT TKNVLSNKKVATTEQAKAEVKNPQNTQISGTQYIQDPAIVQPKTANNNKNAQVSQK VDTAQVNGDTRFANOSATNTNNTQPVAKSTSTPAKTNNTNVTNAGYSLVDDDDNSEHQI NPELIKSAAKPAALETOYKAAAPKAKTEATPKVTTFASAQPSVAATPKTSLPKYP |
| gene | /gene="SAV1054" complement(64696. .65130) /gene="SAV1054" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB57216.1" /db_xref="GI:14246824" /translation="MFSKVNQKMLEDCFYIRKKVFVEEQVPESEIDEVESESIHL IGYDNGQPVATARIRPINETTVKTIERVAVMKSHRGQGMRLMQAVESLAKEGFYVA TMNAQCHAIPEYSELNFKMGNIFIEEGIEHIEMTKKLTSLN" complement(65285. .65755) /gene="SAV1055" complement(65285. .65755) /gene="SAV1055" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB57217.1" /db_xref="GI:14246825" /translation="MSRTEKIANINGMFMNLEQQIIHSQDMAHFSEFFYVNVHEHR ENYBALLYKNSIDNPIDVGCAYILALPEIFNSVDFSELPFSWYDENGITETMK SLSPLOYLAALAEVDVNIFFKPSGFTGMNMNNNIAQMRIFWQYTAIRKEAL" complement(65803. .67020) /gene="SAV1056" complement(65803. .67020) /gene="SAV1056" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB57218.1" /db_xref="GI:14246826" /translation="MNKFLKYLILLALVLIVPVPATLLFKTSQDAFESSQDSKNA NRQNLRDKNVNPBEQPIISILFLGIDDNGRRKKGQDAEHSRSDAMILITTFNQSKHQI RMLSIPRDTISYIPKGYIDKITHAHAYGPIAAMDSVEATMNPVDPYVVRNNKAFY EAVNELGGIYDVPYDLNPNEDTDGKIKIKKGQKLGDEALAVARTRHHSDSLKRQ QRQNELIKILFOKAQEVDSIDKLDNVIQIVGNKAKHNLTNSEIKALAKMYLTNDVEIK TAQLKGGDKMLNGIYYHPSPVESIQYANLLRKOLESPIINDKNDLDOVDORVNHGSL IPLTPDLSSLLURKEQNDTDDKDKTSNENNDSNDSNSQOQPATDQMLNQNQSGTQQ APOASNQNGYVN" 67471. .68664 /gene="fmt" 67471. .68664 /gene="fmt" /note="SAV1057" /codon_start=1 /transl_table=11 /product="autolysis and methicillin resistant-related protein" /protein_id="BAB57219.1" /db_xref="GI:14246827" |
| CDS | /translation="MKPNKVLVIHACVLLFIIISIALIFHRLQTKTHSIDPIHKETK LSDNEKYLVDNRKKEKVPKSLKVEYNSKDPKRYKKIDKYLQSSLFNGSVAIYENGKLKM SKGYQODFEKGIKNTPMFLIGSAOKFSTGILLKOLEECHKININDPVSKYLPWFK TSKPIPLKDLMHOSGLYKYSKSDYKNLDOAVRAIOKRGIDPKYKKHMYNDGNLYL LAKVIEVTKSYAENYTTKIGDPLKLOHTAFYEQDFPKYLAKGAYANSTGLSFLRP NILDQIYAGNLNPTDMDGKLITQIQYKLFSPKTNPLPLEHFTQYDPEYRIGFY AKPTLNKNGLGGFGQVFTVYVYNDKYVVVVVYVVALNVLNKGNNVEIRIKHTYNDILKQNKPYNTK GVIVQ" |

gene

complement(69201. .69491)

/gene="SAV1058"

CDS

complement(69201. .69491)

/gene="SAV1058"

/note="similar to quinol oxidase polypeptide IV QoxD"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57220.1"

/db_xref="GI:14246828"

/translation="MSTIMKHTVGFASIVLTLLAVVTVLTSLTFHAKLTIIEGFAP

VQAGLQLLMPHLTEGDKRLQTEPKVFALVITLCFVVGVYWMQGGHSHL"

complement(69488. .70093)

/gene="goxB"

complement(69488. .70093)

/gene="goxB"

/note="SAV1059"

/codon_start=1

/transl_table=11

/product="Quinol oxidase polypeptide III"

/protein_id="BAB57221.1"

/db_xref="GI:14246829"

/translation="MSHDTMTIDSRTHGELNKLGFWIFITAEPALEFGLTFATLLTLQ

HGGDYAKMTTELEFELPLVLMFALLPSSYTCGAIYYMRQEKQKLMFMWILLTL

GLVPVGEIYEFAHAYSEGYNPTIGYSWSSFFILLGTHGCHVSIGIVWAICLLIQIOR

RGDKYNAPKLFIVSLYWHFLDVVWFVIFTAVYMIGMVSS"

complement(70083. .72071)

/gene="goxB"

complement(70083. .72071)

/gene="goxB"

/note="SAV1060"

/codon_start=1

/transl_table=11

/product="Quinol oxidase polypeptide I"

/protein_id="BAB57222.1"

/db_xref="GI:14246830"

/translation="MNFPPDQLLVKGNMIMTAQIGAPFLVIGLIAVITYFKLWKYLY

KENFTSYDMKIGIMYLICAVMEVRGIDALIRLAQLTVPDNKFLSESNHYEIFFSTH

GVNFIEMAMPFIFGLWNIIVPQIGARDVAFVPLNNVSFWLFPAGMILFNLSPIIGG

SPAAGNTYAPLAGEFPGVGNVYLLAIOISGLGLTATGNFVFTILRCKTPTMKFM

QMPMTFTTITLIVLAPPLTVLALMTDRIFDTAFETVAHGGMPMLWNFVW

KGHPVYIVLPARGIYSEIIPFARKLRGHSQWVATAGIAPLSLVWVHHFFTMG

NGALNSFFSLIGITPGKLFNLWLLIYIGRITFSPMLFSLAFIPNLLGGVT

GVLAMASADYQHYNTYFLVAHFYTLTVGVFACLAGLIFWPKMGYKLNELTNK

CFWFFMIGFVNCFLPQFILGDGMPRLTYMPSDGFLLNLISTIGALLMAIGFLF

VSVTVSHFSPREATGDNMDGLGRTLEWTTASAPPKYNAITPDWNDYDFVDMKE

HGRYLDHNNYKDIHPMNPVPGVIGIEMTIGGFRLIFETVIPALICLGFICTMIY

RSQIDHGXYHPAAEVAETEARLEARIKREAVSHES"

complement(72071. .73171)

/gene="SAV1061"

complement(72071. .73171)

/gene="SAV1061"

/note="similar to quinol oxidase polypeptide II QoxA"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57223.1"

/db_xref="GI:14246831"

/translation="MSKFKSLLLFGPLLSCSNIEIFNAKGPVASSOKFLIYSI

VEMLVICFVVLGMFAIFKYYSYNKNAESGMHNAIETIWFVPIIIVAAIAIPV

KTLDYDKPKSEKDPVNVYASGAKWFFAYPDEHETVNTLTIPKDRPVFKLQAM

QDWTSPVQLGGQKYAMTGMWNTLEASQTGTFRGNSFNNGEFSRQTFFKNVYS

KTYDKWVKGKTKLDDQTFDKQLLSPNKALEFNGTHMAFVDPADPEYIFVAY

KRFNFKLDPNFTSEENMKFVSDKPLIPARKAQITNANYKRHKMGLMILGNDEPYN

EFKDKSNKAKEMKISKDAQDQDNDHGGGH"

complement(73742. .74059)

/gene="SAV1062"

complement(73742. .74059)

/gene="SAV1062"

/note="similar to chitinase B"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

gene

complement(74923. .75783)

/gene="fold"

complement(74923. .75783)

/gene="SAV1063"

/note="fold"

/codon_start=1

/transl_table=11

/product="Fold bifunctional protein"

/protein_id="BAB57225.1"

/db_xref="GI:14246833"

/translation="WVAKILDGKOIAKDYROGLODOVEALKEKGTPKLSVLVGNDC

ASQSVRSKKKAEEKICNISEIVHLEETATEEVEVLNLRNLNDDSDSGILVQVPLPK

QVSEKQKILEAINPEKDVDFHPINIGRLYIDEQTFVPTPLGIMEIILKHADIDLEAK

AVVIGRSHVGPQVSKLLQKNASVTLIHSRSDMASYLKADAVIVSAVGKPSLVTKD

VVEKGAVIDVGNTPDENGLKGDVDYDAVKEIAGAITPPVPGVGPLITITWVNLNTLL

AEKMRRGIDS"

75984. .76466

/gene="SAV1064"

75984. .76466

/gene="SAV1064"

/note="similar to phosphoribosylaminoimidazole carboxylase

PurE"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57226.1"

/db_xref="GI:14246834"

/translation="MKVAVIMGSSSDMKIMOESCNMLDYPEIPYEKOVVSAHRTPKMM

VOFASERERGINIIAGAGAAHLPGMVASLTTLVPGVPIETKSLGIDSLISIVO

MPGPIVATTAIGAAGAKNAGILAAARMLSIQNSLSLVEKLNQYESLSIQKVEDMQNELQ

76453. .77577

/gene="purK"

76453. .77577

/gene="purK"

/note="SAV1065"

/codon_start=1

/transl_table=11

/product="phosphoribosylaminoimidazole carboxylase carbon

dioxide-fixation chain PurK homolog"

/protein_id="BAB57227.1"

/db_xref="GI:14246835"

/translation="MNFENKLFGATIGIIGGQGLKMKMAQSAQRMGYKVVLDPSEDC

PCRYVAHEFIQARYDDEKALNQLGQKCDVITYEFENISAQOLKLLCEKYNIPQGYQAI

QLQDRLTERETLKSAGTKVVPFISVKESTDIDKAIETLGYPTIVKTRFGYDGGOV

LINNEKDLQEGFKLIETSECVAEKYLNTKKESVLSLTRGNNNQITFFPQEGNEHRNOI

LFTITVPARIDKTAEEKEQVKNIIQSIHFITGTVEFFIDSNNOLYVNEIAPRPHNSG

HYSTACDYSQDFPHILAVTQSLPNSLSLELLKPAVMNNLLGKDLLLENEFNEHPEWH

LHYGKSGRSDSRKMGHMTVLINDVNOTEQDMYAKFECSN"

77381. .78285

/gene="purC"

77581. .78285

/gene="purC"

/note="SAV1066"

/codon_start=1

/transl_table=11

/product="phosphoribosylaminoimidazolesuccinocarboxamide

synthetase homolog"

/protein_id="BAB57228.1"

/db_xref="GI:14246836"

/translation="MTLLYEGKAKRIFSTNQENELRVEYKDEVTAGNKAOKDTWAGKG

RLANQTSIIFPKYQENGIESHFIKQISETQLVKPKKIIIPLEVVVNRNIASGSIKKRL

GFNGEYFRELPEFFYKNDALNDPLTDDHVKLLINIASDEIDILKSKALKINNVLK

QLMDAMNKLKLVDFEKIEFGKTETGQIILLADEISPTCRIMDKATNANFNKDVYRNTGS

LIETYQIFLAKLEDLK"

78285. .78548

/gene="SAV1067"

78285. .78548

/gene="SAV1067"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57229.1"
/db_xref="GI:14246837"
/translation="MKTIELHTLPQVLDLTQQTITRAVHDLGYAQVNDIRVGKVLV
MIVDEVSDEKVVHNIITLSEKLFANTVIBEYSYKVLDDKEKA "
78550. .79221
/gene="purQ"
78550. .79221
/gene="purQ"
/note="SAV1068"
/codon_start=1
/transl_table=11
/product="phosphoribosylformylglycinamide synthase I"
/protein_id="BAB57230.1"
/db_xref="GI:14246838"
/translation="MKFAVLVFPGSNCDRDMFNAAIKSGVEAPYVDYRETSLSGFDG
LIPGFSFGDYLRSGAMASVAFIISVKRLATEGKPVLGVCNGFQILTEIGLLPGALL
HNDSHLFI SRNEELETIVNNQTAFTNLYEQGEKVIYPVAHGEGHYCTDEIYQOLKANN
QIILKVVNPNNGSYDDIAGIVNEKGNVCGMPPHPERALETLLGTDSGVKLF EAMVKWS
RQHV"
79214. .81403
/gene="purL"
79214. .81403
/gene="purL"
/note="SAV1069"
/codon_start=1
/transl_table=11
/product="phosphoribosylformylglycinamide synthetase"
/protein_id="BAB57231.1"
/db_xref="GI:14246839"
/translation="MSKFTEPSVEETIKKVVQDMGLSDQVEKVKCDILGRQPNFTET
GIFSVMSEHCYKXSKPLKQPTSGEHVLMGPGEGAGVLDGNQAVFVKVESNNH
PAIEPYQAATAGVGGIIRDIIVSIGARPINLLNRFGEIDNKNQRKGLVVKVGIGG
YNGICIGTAGEIEFEDRYDGNPLVNMVCVGINHDMIOKGTAKGVGNSVIYVGLKT
GRGHTGATFASBELTEESSEKRPVQIGDPFVGKLMLEATLEAIFDLVLVQDMGA
AGITSSSMAAKGGSLHRLKEQVTPREGIPSPYEMMLESETQERMLLVVEKNEQKF
LDLFKHELDASVIGBVTOTNRFLTYDDEVADIPEVLADAPVY ILEGEKDYNT
SKNDYTHIDVDTFFKLLKHPITASKHLYDOYDOOVGANTIIKPGLASVYRVEGNT
KAIASTIDGEARYVNNPYEGGKMVVAEAYRIALVAGATPLAMTDCLTGSPPEKKEIY
QQLIDSTKMAEACDILKTPVSGNSVSLYNETKGTSIFFTPVVMGGLIENVYLNDF
EPQGDKLYLIGTDKDDFGSQLEKLYIGKVNHEFESLDLSSEVEKGESIKTAIREGL
LSHVQVVGKGLLITLAKLSAHYGLGKSIDITNAOLFSETQGRVYVSVKSKGTENI
DNAIEGLTDSDNFKVTPPYTEISENSVSDIKQIWEGAIAQLTITQD "
81382. .82866
/gene="purF"
81382. .82866
/gene="purF"
/note="SAV1070"
/codon_start=1
/transl_table=11
/product="phosphoribosylpyrophosphate amidotransferase"
/protein_id="BAB57232.1"
/db_xref="GI:14246840"
/translation="MPNYSGLNEECGVFIWNHP EAAQLTYMGLSHSLQHRGQBGAGIV
VSDNELKGERGLLLEATINDOMERLKGYQHAIGHVRYATSGNKGIOQPLFHF
YMSVGICHNGNLINAKSLRONLEKQGAIFPHSSDDETEVIMHLIRRSKAPTFEALKES
LRRVKGGFTHAILTDKALYGAVDPNAIRPLVGMKMDGTYYILASETCAIDVLGAEFVQ
DIHAGYVITNDKGTIVKSYTHHTTTAISAMEYIYFARPDSITIAGRNVHAVRKASGK
LAQESPVNADWIGVPNSLSAASGYAEETGLPYEMGLVKNOYVARTFOPTQOLREQ
GVRKLSAVKDIVDGKNIILVDDSIVRGTTIRIRVKMLKDSGANKVHVRIASPEEFEP
SEYGDVSTTAELISAKSPEELKDYIGADSLAYLSVDGLIESIGLDYDAPYSGLCVE
STFGDYPAGLYDEANYKAHLSHRQKYISKNKHFDFSGENLNV "
82859. .83887
/gene="purM"
82859. .83887
/gene="purM"
/note="SAV1071"
/codon_start=1
/transl_table=11
/product="phosphoribosylformylglycinamide cyclo-ligase"

/protein_id="BAB57233.1"
/db_xref="GI:14246841"
/translation="MSKAYEQSGVNIHAGYEAVERSMSSHVKRTMRKEVIGGLGGFGAT
FDLSQLNMTAPVLVSGTGDGVTKLKLAIDYKDKHDSIGIDAVAMCVNDILTTGAEPLYF
LDYIATNKVYVEVIEQIVKGISDACVETNTALIGGETAEMGEYHGEEDYVAGFAVGA
VEKDDIVDSEKREGQVIGLASSGHSNGISLVRLKINESGIDLASINRRPFDIVF
LEPTKLYIPKVPALKKREVSIKAMNHITGGGFYENIPRALPAGYAARIDTTSFPTPKIF
DWIQOQGNIDTNEMYNIFNMNIGYTVIVDEKDSRALKILAEQNVEAYQIGHIVKNES
TAIELGV "
83890. .84456
/gene="purN"
83890. .84456
/gene="purN"
/note="SAV1072"
/codon_start=1
/transl_table=11
/product="phosphoribosylglycinamide formyltransferase"
/protein_id="BAB57234.1"
/db_xref="GI:14246842"
/translation="MVKIAIFASGSGSNFENIVEHVESGKLENI EVTALYTDHQNAFC
IDRAKKHDPVYINEPQFDSKAAYEQHLVTLNKKOKVEWIILAGYMRLIGPDLLASF
EGKILNHPSLPKYKIDAIGOAYHSGDPTITGSTVHYVDCGMDTG EII EQRQCDIRP
DSSKEQLEKVKKLEYELPSVIAKIV "
84468. .85949
/gene="purH"
84468. .85949
/gene="purH"
/note="SAV1073"
/codon_start=1
/transl_table=11
/product="bifunctional purine biosynthesis protein"
/protein_id="BAB57235.1"
/db_xref="GI:14246843"
/translation="MMKAILSVSNKGTIVEFAKALTQLNVELYSTGCTKRILDEANV
PVRVSDLTFPEIMDGRVKTLLHPAVHGGILADBNRPQHNLSELSEQHLIDLDMVVNL
YPQQTVANPDVTMDAEINIDIGPTMLRAAKNTAKHTVTIIVHPADYHEVLTRLRND
SLDESYSQMSIKIVFEHTAEYDEAIVRFKPDKETURYGENPOOSAYFVRTSNAKHTY
AGAKLHGQLSYNNIKDADATLALVKFTDPAAVAKHNPBGCVGIGDTIEQAFOHA
YEADQSIFGCI VALNRVTPELAEOLHSIFLEVIITAPKFTDEALDILKOKKNVRLJE
IDMTIDSNEEFVYSGGYLVQDKNVVPKEEMKVVEVATPDEOWEAMLLGKWKVP
SVKSNAILLSNNKQTVGIGAGOMNRVCAKIALERAIEINDHVALYSDGFFPMGDTVE
LAHQHGKAIQPGGSIKQDSIDMANKHGIAWVVTGTRHFKH "
86079. .87218
/gene="purD"
86079. .87218
/gene="purD"
/note="SAV1074"
/codon_start=1
/transl_table=11
/product="phosphoribosylamine-glycine ligase"
/protein_id="BAB57236.1"
/db_xref="GI:14246844"
/translation="MTPIAEVHTEISESNHQGILDFAKQONVDVNVVIGPPEQLIDGLA
DILRANGKVFVGNKQAAQIEGSKLFAKTKMKYINPTADYKEVERKKDALTYIENCE
LPVYVKKDLGACGVIIADTIEAKSAIEIMYGDEEGTVVFTELEGEFSMLTFV
NGDLVPFDCIAQDHKRAFDPDEGNTGMAVCPVPHISDDVLKLTNETIAQPIAKA
MLNEGQFVGLYIGAILTLTKDGPVTEFNARFGDPEAOVLLSRMESLMQHOHIDLDBG
KRTPEFKNESIVGVMLASKGYPDAVEKHGKVSQFDLNEYFVSGLKKQGDFTVTSGG
RVILAIGKQNSIVQAOQADATKEVQSQSDHLFTRHDTANKALQK "
complement(87486. .88292)
/gene="SAV1075"
complement(87486. .88292)
/gene="SAV1075"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57237.1"
/db_xref="GI:14246845"
/translation="MYEQWKVROSFDVDVNIITKFLGIILFFFIIFVHNFDFMLYIT
ILMLVFLLMNGTQFKITAIFITISIGFALLSSFLMILYNGKSHVILKFGFIHTTES
LVRLGHVSLRFTIAISFFGILITATSQVVMIFYSLMQHLKYGKSVAVAFMAAIRMILPM
ISSLIQRLSRKMYQIDAANYRGLKRFRKHLIIPILSONIRKAHQLSVAMESKGFSD
GPTYVYVPFSYKDIIFIIVLIAITITMAYYCSITLPIITGIDDVRLGRVG "

```

gene      complement(88285.. .89685)
CDS       /gene="SAV1076"
          complement(88285.. .89685)
          /gene="SAV1076"
          /note="similar to cation ABC transporter"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="BAB57238.1"
          /db_xref="GI:14246846"
          /translation="MLKVSDLRLKYPNGQRKIFDHLNITIQDKREKVLILLPGSGCGKST
          LNLVSGIVPNLIELPMKVDYDELIVDPLSGVIFQDPDSQFCMPKPYEELAFVLNROVP
          RDMADALIINALNVLNVTPTYIKDLSGGMQKLAIVETILQOSKTLFLDEPTAML
          DQATEDLWTKLLELREDQTVLVEKHVHWHVDRVILMDYNGNIIADECEFEILQ
          KYVHLLSEGVWHPRAWEPAPSRVDPPTNSHLLQFNKRIIRGKSTLLSFSLEILQ
          GENITITGANGSKRTILLESIMOLIKIYQGVYFENQRLTKIKHAAKHMVLYVQNPELQ
          FINTSVYDEINHFNLKSDQDDETIQLLLDLQNVKDOHPYELSIGQKRRLSVAT
          ALSKADIIFLDEPTFCLDSSHNTFQIKLQFQRINLQSGIVMVTTHDEILERYPSRRL
          KISDGAALLDCDGTNV"
          complement(89700.. .90275)
          /gene="SAV1077"
          complement(89700.. .90275)
          /gene="SAV1077"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="BAB57239.1"
          /db_xref="GI:14246847"
          /translation="MSKGLKLSILVTVLISVVFAVIYNLMFVYNGVQATGHLRLQ
          TNGVFWMAIVCYLIIKPKGIALLAFAAGAGETIIMGRFDIPTIVAFYAFQGLACELV
          FAIFKYQSRVVMVLAAGCTAIAAFPIDYFYGYLNEVAGWNLTFLVFRLLISGAVVA
          GVLSYLLVRLADTKGTVKLFRPAAKEDYNL"
          complement(90656.. .91105)
          /gene="SAV1078"
          complement(90656.. .91105)
          /gene="SAV1078"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="BAB57240.1"
          /db_xref="GI:14246848"
          /translation="MPLQLLVVRGYACARRMCSSSSLQLLVVRGYACARRMYPST
          LQLLVVRGYACARRMYPSSLSQPLVVRGFFACARCMYPSFTLQLLVVRGFFACARRM
          YPSFTLQLLVVRGFFACARRMYPSTLQTIIRKPLCTGSSLLKKRYD"
          complement(91111.. .91245)
          /gene="SAV1079"
          complement(91111.. .91245)
          /gene="SAV1079"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="BAB57241.1"
          /db_xref="GI:14246849"
          /translation="MSNENQNKAKEAKEVEEKLKDKKEEKTEDINOTKQDIQDTLN"
          91396.. .92691
          /gene="SAV1080"
          91396.. .92691
          /gene="SAV1080"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="BAB57242.1"
          /db_xref="GI:14246850"
          /translation="MSFLRKHAEIFSYLIGVLSFTGLIILINPLIKQLNGRKKVD
          THVNWFLNAFFSEIIKVMRSFNGFPPIVSAIVIIIFGILVMLIGHTLRTIKYD
          DISFFPLGIMYFIITLMTQVYGFPAVIFIPPTIHIGYVYKDELNOENVKNHF
          MWIIVSYGSLYTOIALYGRIDANEIESIDILSVNAFFIIMWLLGQMAIWNFLRR
          ALPLTKQLGEPELPSRTSGKNTNQT'KHLQLODKTTEYAKTRFSVDLKIIRAK
          RDKFKKYNDAIIDQEDDIPWMRKPKWKMVLEFCGVFIPLTFLEFNRRNALFY
          SGDWKLSQVIVIEWVTLILLFIIIAVIATTLTFLHKGKRYFYQLQFMGSLIFLKKLLT
          EFINIHGHLGLLSVFEITPTLLMLLAIISIYSLQLRERP"
          93116.. .94288

```

```

CDS       /gene="SAV1081"
          93116.. .94288
          /gene="SAV1081"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="BAB57243.1"
          /db_xref="GI:14246851"
          /translation="MKIATLNKGETKYFNGYPLIEEDIYSODHLKEGDIQIVTQK
          SQVATAYVGRQKHGLGWLTYDKAORINTAFVFKLFTALAEEDYFENIDGTNAFLR
          ENAGDGVGLTIDNYDGHLLIOWYSGKIYKFKYAILAEVRKYFDYKSIYEKVRFKDS
          EYSGGVEGAPPEPPIVENFTFYNVDELGLMTGIFLQOKEVRKKLRQYAKARHV
          LNFSTYGAFAAASEASTTSVDIANRSRSTENFNLNADPQSOYIYVMDTFDF
          YKTAARHGSYDTIVDPPSPFARKRTFSQDKYDKLINGALNLSSEGTLLICTNA
          SVYPLQKQNTIKTKITLESQVDYELTEVMGLPKDFKTHPHYKPSKYLKAVFVNIRH"
          94342.. .94884
          /gene="SAV1082"
          94342.. .94884
          /gene="SAV1082"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="BAB57244.1"
          /db_xref="GI:14246852"
          /translation="MGKKNLTSLNKNIGNSVFKIENVDGKGAMPPTIQELERROR
          AEAIVRKSLMSSTMSVPIPGLDGVDLKLMDKIIEDNKNYIGLDHKQVNSLGDDYK
          ERIMSAAIQGSQFIGKRISSAFLKIVIRDVAKRTAAKQTKWFPVVQAVSASISYFF
          MNKIKDHIQKCNVKNVM"
          95038.. .95304
          /gene="PfsH"
          95038.. .95304
          /gene="PfsH"
          /note="SAV1083"
          phosphohistidin-containing protein"
          /codon_start=1
          /transl_table=11
          /product="phosphocarrier protein Hpr"
          /protein_id="BAB57245.1"
          /db_xref="GI:14246853"
          /translation="MEQNSYVIDETGIHARPATMLVQTASKFDSIQLENGYKKNL
          KSIIMGVNSIGVGKDAEITTIYADGSDSDAIIQAISDVLSEKGLTK"
          95307.. .97025
          /gene="Pts1"
          95307.. .97025
          /gene="Pts1"
          /note="SAV1084"
          /codon_start=1
          /transl_table=11
          /product="phosphoenolpyruvate-protein phosphatase"
          /protein_id="BAB57246.1"
          /db_xref="GI:14246854"
          /translation="MSKLIKGAASDGAIAKAYLLIVEPDLTFDKNEKVTDEGEVAK
          FNSAIEASKVELTKIRNNAEVLQGAIAIFDAHLHLLDPDLIOPIDQK IKNNENANA
          ATALDVTTFQVYTFESMDNEYMKEAADIROVSKRVLSHILGVELPNSMIDESVVI
          VGNDLTSPDQAOLNKEFVOGFAINIGGRTSFAKKELEIFAIYKTSITQEVKQGD
          IIVDGLNGDVIVNPTEDELIAYQDKRECYFADKKELEKQLRDADTVTDGVHAEALAN
          IGTNDPLGVIENAGQGLYFTEFLYMRDQMPTEEQEYAYKEVLEAMDGKRVVVR
          TLDIGDKELSYLNLPEMNPFLGYRAIRLCAQODIFRQLRALLRASVYKGLNMF
          PMVATINEFREAKAILLEKENLKNEGHDISSDIBGINVEIPATAALADVFAKEVDF
          FSGITNDLIQYTLAADRMSERYLYQYPNPSILRLVKQVIEASHKGEKWTGCMGEMA
          GDETAIPLLGLGLDEFSNATSIILKARQINGLSKNEMTELANRAVDCATQOEVEIEL
          VNNYVK"
          complement(97262.. .97495)
          /gene="SAV1085"
          complement(97262.. .97495)
          /gene="SAV1085"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="BAB57247.1"
          /db_xref="GI:14246855"
          /translation="NSEIIVYTONDCPCTPVKYNVNEHQIDFEERNINNOQYRNEMI
          DFDAPSTPILLNGPNMHVDLDEINKVLNIQD"

```

gene

CDS
97693..99054
/gene="SAV1086"

97693..99054
/gene="SAV1086"

/codon_start=1

/transl_table=11

/product="cytochrome D ubiquinol oxidase subunit 1 homolog"

/protein_id="BAB57248.1"

/db_xref="GI:14246856"

/translation="MDTVEISRELTAMTAVHIIIPATIGVGMPLMFAETAEFLGIRKND
LOYIMARKWAKAYITIVAVGVVGTMTIIGLQLSLIWPFMEMGGHVIAPLPMETTFAP
EFEAFLSIYITWDRFNKNTWTHFLISIPVIGGSFAFFITSVNSFMTTAPGPELKN
GKMVNPQIEAMFNPSFIVRSFHVITTAGMTMAFYIASIAAFKLRNRPKDTVYHKK
ALKSMIVGFFSTLLSMLAGLSAKFLHKFQPEKLAAYEWHFDTSSHAKLLIFGVLDCE
KTOQVKAELPLGLLSFLADNSVTKVOGLNDFPKSLHPMTIVHVFFDLMTWMTGLCF
VISGVVLTLPFKLKRKESTHKWMLYGLITGCPASMLAIEFCWFLTEMGROPPIVRGY
MRVAAATQAAGITFVTILFGILYIILMYTCAVYDIRMKNKPATEDVNRKAKOGGE
IEK"

99051..100070

/gene="SAV1087"

99051..100070

/gene="SAV1087"

/codon_start=1

/transl_table=11

/product="cytochrome D ubiquinol oxidase subunit II homolog"

/protein_id="BAB57249.1"

/db_xref="GI:14246857"

/translation="MIYAFIGITVLWLFPCYIIIASIDFGAGFPALHSKLTGDEKKI
NHLISRYLNPWEVTNFFVFFGVGFPPESIKYLTGLVLLPVGSIALIMISLRNSF
YAFENYGDTLWAMIMYGVSLIPAGLSALTALITTEGGYINVRNNVLDLDWVLLLS
PFANVSFLAIIISVLYISSGFLTYVAKKANDEPAVNLTRQMHILFGPPMIIICLFLVFL
SLRIONSSEHYSAVEDYWNMFASIFLFPALASLLTFPFFKKKGLAFVFFVLQWMEAFPG
YGISKLPLLLFPFKITDAYVNPMEGWTIVIVFIILGILLLLPSLLILLLRLEFVEDKEYV
EGKKS"

100203..100865

/gene="SAV1088"

100203..100865

/gene="SAV1088"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57250.1"

/db_xref="GI:14246858"

/translation="MKEYVVIIGLRCFGGSIVRELNALMDVMYIDHDENRVNYESDI
ATHAVADTDEAVNKSLSGIRNFDHVIIVAIGENIOSSTLTLLILKELGVKKVTKAQON
DYHAKILNKIGADTVYHPERDMRRRIAHNVASASVLDYLELADEHSIVELKATERMAG
QSIIDLDRAQYGINIIAIKRGKEFIISPNNPINLEIGDILMIGHDNDLNRFEKNIA
TR"

complement(101111..102808)

/gene="SAV1089"

complement(101111..102808)

/gene="SAV1089"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57251.1"

/db_xref="GI:14246859"

/translation="MKQLHPNEGVYVALGGIGEIGNKYAYVEYKDEIVIIDAGIKFPD
DNLGIDVIVPDYTVLQVQDKIVGLFITHGHEDHIGGVFPLLKOLNIPVIYGGPLALG
LIRNKEEHLRLTAKLAINEDSVIKSKHFTISFLTTHSPITGYVDTPEGVKV
HTGDPKDFPTVPGKPAKIAQALGEEGVCLLSDSTNSLYPDPFTLSREYGVONDKI
FRNCKGRILFATFASNIYRVQAVEAAIKNNKVIYFGRSMENNIKIGWELGYIKAPP
EFFIENKINTVPKHELLILCTQSOGEPMAALSRIANGTHQKIKIIPEDTVVFSSEPI
PGNTKSIAMTINSLYKAGADVIHSKINSHTSGHSGQDQQLMLRLIKPKYFLPHGE
YMLKHAHGETGECQVEEDNVFIDIGDVLALTSDSARKAGRIPSGNLIWDSGICDI
GNVTRDKLLESEGVLIWVWSIDINTKLLSGDPIISRGFVYMRSGOLIVDAORKI
KTDVLSKLNQNDIQWHQIKSIIETLQYLPFEKTARKPMILPVIKMYNEQKESNNK"

complement(102786..102980)

/gene="SAV1090"

complement(102786..102980)

/gene="SAV1090"

gene

CDS
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

/protein_id="BAB57252.1"

/db_xref="GI:14246860"

/translation="MCGENTQSLYVEAQTEQVRRYLKDRNFNIEFIITKLEGAHLIDYE
KETQNTLMWRLNNETITSK"

complement(103508..104059)

/gene="pdf1"

complement(103508..104059)

/gene="pdf1"

/note="SAV1091"

/codon_start=1

/transl_table=11

/product="formylmethionine deformylase homolog"

/protein_id="BAB57253.1"

/db_xref="GI:14246861"

/translation="MLTWKDIIRDGHPTLRQKAAEELPLTKEKETLIAMREFLVNS
QDEETAKRYGLRSGVGLAAPQINISKRMIADVLPDDGSGKSYDMLVNPKITVSHSVQE
AYLPTGEGCLSVDDNVAGLVHRHNRITTIKAKDIEGNDIQLRLKGYPAIVFQHEIDHLN
GVMFYDHIDKKNHPLQPHPTDAVEV"

104424..105050

/gene="SAV1092"

104424..105050

/gene="SAV1092"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57254.1"

/db_xref="GI:14246862"

/translation="MKFGKTIADVILASSVLLAGCTTDDKKEIKAYLKQVDKIKDDDEPI
KTVGKIAELDEKKKLLKEDVNSKDTAVRGKAVKDLIKNADRLKEFEKEEDAIKKSE
QDFKAKSHVDINDNDVKKKEVKQLDDVLKKEYKLHGEVAKAYKAVANSEKTLFKYLN
QNDATQQGVNEKSKAIEQNYKKLKEVSDKYTKVLNKVCKEKQDVDFK"

105221..106333

/gene="pdhA"

105221..106333

/gene="pdhA"

/note="SAV1093"

/codon_start=1

/transl_table=11

/product="pyruvate dehydrogenase E1 component alpha subunit"

/protein_id="BAB57255.1"

/db_xref="GI:14246863"

/translation="WAPKLOAQFDVAKVNLNDTQSKFEMVQIILDENGNNVNDLVPLDT
DQLVELMERVMVTRILDQRSISLRQGRKIGFYPTAGAEASQLASQYALEKEDILP
GYRDPYQIWIHGLPLTEAFLSRGHFKGFPPEGNALSPQIIIGAQYIAAGVAFAL
KRGKNVAITVTDGSGSQDGYEGINFAAAAYKAPAFVIONNNYAIISTPRSKOTAA
ETLAKOKATAVGIPGIQVDMDALAVQATKEARDRAVAGEGPTLIETTYRYGPHITMA
GDDPTRYTSDDEAEKKDPLVRFRKFLKGLNEDKENEVIERAKADIKAAIKEA
DNTEKQTVTSLMEIMYEDMPQNLAEQYEIYKEKESK"

106337..107314

/gene="phdB"

106337..107314

/gene="phdB"

/note="SAV1094"

/codon_start=1

/transl_table=11

/product="pyruvate dehydrogenase E1 component beta subunit"

/protein_id="BAB57256.1"

/db_xref="GI:14246864"

/translation="WQMTWQAINDALKTLEKNDQDVLICGEDYGVNGGVPRVTEGL
QKEGEDRVDPDTPLAESIGGLGAMLAGEFGRPVMEVQFLGFVFEVFDRIAGQIARTR
FRSGGTDFVPTIRSPFGGVHTPELHADNLEGLKAVPGLKVVIPSGPYDAKGLLIS
STRSDPVVYLEHMKLYRSFEEVPEEYITDIGNKVKEGNDISITITYGAMVQESM
KAAEELKDGYSVEIDLTQVQIDVDTIVASVEKTGRAVVVQEAQRQAGVGAAVAAE
LSERALLSLEAPIGRVAADTIYPTQAEVNVNLPNKNKDIIEKAKETLEF"

107405..108697

/gene="pdhC"

107405..108697

/gene="pdhC"

gene

CDS


```
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57266.1"
/db_xref="GI:14246874"
/translation="MNRKPELIMANTANSISIIYLLVMGLSYFSLKSGNASQREELAK
QLSQNGNVSLMDLQTTMGALAIILISYLGIFATICIKGRKKLSIILFVIAIVSL
MALNLIAIWLWIVIMLISLKSKESTHKDDIYH"
complement(116892..118244)
/gene="SAV1105"
CDS
complement(116892..118244)
/gene="SAV1105"
/notes="similar to Mn2+-transport protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57267.1"
/db_xref="GI:14246875"
/translation="MNNKRHSTNEQLSDEINNTIKFDRHSNNKQKFLSFLGGLLVA
VGYMDPGNWTSMQGAQGYTLLEVLIVLSSAMLLQSMYVRLGIATGMDLAQMTRH
YLSRPIAIFWIAELAIATDAIEVIGSAIALNLLFNILPLIVGALITVDVFLLLFI
MRYGFRKIAIVGTFTFTVLFIFIEVYISSPOLNAVLNGFIPHSIITNNGILYAL
GIIGATIMPHNLYLHSSIIVQSRTYSRHNNEKAQAIKEATIDSNIOLSIAFVVCILL
VLGASLFFNSADDLGGFYDLYHALKTEPVLGATMGAINSTLFAVALLASQNSITIG
TLAQIIVMGFLURHIPNWLRLITRSLAIVIPVCLIIIFKGNAAKIEOLLVFSQVFL
"
/transl_table=11
complement(118428..119042)
/gene="SAV1106"
CDS
complement(118428..119042)
/gene="SAV1106"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57268.1"
/db_xref="GI:14246876"
/translation="MTKYTFKPKDFKAFNVEGLDARMEALNEYIRQLHELGEYFSDF
FTSQGETPYHVAKHARRSVNPKDTWAFATSKRGYKMLPHFQIGMFEDQLFVNF
INHEAKDRAKFAVERKEFAIQQLPDDYRVCLIDHMKPKPKFKOLTDDDLKEAIO
INVKGEFTIARAITPQDKRLKSDRAFIATLEETFDQLPEYSA"
119199..120026
/gene="SAV1107"
CDS
119199..120026
/gene="SAV1107"
/codon_start=1
/transl_table=11
/product="myo-inositol-1(or 4)-monophosphatase homologue"
/protein_id="BAB57269.1"
/db_xref="GI:14246877"
/translation="MALYGAQGLIQEAGIRIKQLMEQNLTITETKSNPNDLVTVNDKA
TEDFIDFTILETPNHQVLGEEGHGHDIDTSKGTVMVVDPIDGTNLFVHQENFAISI
GIYIDGKPVAGFYDVMADVLYHAKVGEGAYRGSQPLKPLNDSNLROSIIGINPNWLT
KPLIGEIFKEIYVNDSSARAYGSAALEIVSVATGNLEAVMTPLQPMDFAGGLVILE
VNGQASNLGEPILTISGPNLSILVGNRGLHQEISNDYLEPHYDTLIQHQEQRKRSK"
complement(120180..120371)
/gene="SAV1108"
CDS
complement(120180..120371)
/gene="SAV1108"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57270.1"
/db_xref="GI:14246878"
/translation="MKQKKSNIFFWFSILAVVFLVLSFVAGSASVNPMLTIFILY
ATFGIGFTTKKKYRENDWL"
120473..122320
/gene="SAV1109"
CDS
120473..122320
/gene="SAV1109"
/codon_start=1
/transl_table=11
/product="GTP-binding elongation factor homologue"
/protein_id="BAB57271.1"
/db_xref="GI:14246879"
```

```
/translation="MTNKKREDVRNIIAIIAHVDHGKTTLVDELKQSGIFRENHVDHER
AMDSNDIERERGITILAKNTAVDYKGTNRINLIDTPGHADFGGEVERIMKVMVGVVLV
DAYEGTDPOTRFLVKALQNLKPVVNVNKKIDKPSRPGCVGVDEVLDLFTLEANEDE
LEFPVYASAVNGTASLDEPKODDNLQSLYETIIDVVPAPIDNSDEPLOFOVALLDYN
DYVIGIGIVERPGKMRVGDVNSLKLDTGVKNFRVTKIEGYGLKRLIEIEEAQAGDL
IYVSGRIGEDINVTGVTTPHDKALPVLRIDEPTLEMTFKVNNSPFAGRGDFVTAQRI
QERLNQOLETDSVLSKNSDSDPTWVAGSLGARKGEMVDMTTDNGLTLRLIFNVPARG
KEIDGYMCEPFFERVOCEVQENAGAVIESLGARKGEMVDMTTDNGLTLRLIFNVPARG
MIGYTFEFSMTRGYGIINHTFEERPRKQAIQIGRRNGALISMDQGSASTYVAILLE
DRGVNMEPGTEVYEGMIVGEHNRENDLVNITTKTKHOTNVRSATQDOTQTMNRPRL
TLEEALQFINDDDELVEVTPESIELRKKILKNVREKEAKRIKQMMQENE"
complement(122460..122627)
/gene="SAV1110"
CDS
complement(122460..122627)
/gene="SAV1110"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57272.1"
/db_xref="GI:14246880"
/translation="MRQVQCIIICDAKVFIDERTTESKRLKNNPDIRTFMCDCKSLRDT
PKORAQHYPLD"
complement(122629..123111)
/gene="SAV1111"
CDS
complement(122629..123111)
/gene="SAV1111"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57273.1"
/db_xref="GI:14246881"
/translation="MIQIKGAVKFPITLDSSTWIFDORKVSIEDLEKGVFDGSKPIDF
EDNHEWNRAILEGOTNPPTLNSEIKYKRAMLEGSPVINMTPEFKHSEPHDSATTIRL
SNEHHDHLDIPDLLPYLFFQFAKDKRRLYSDNGVDSFIYNQEDGYSYRINHVTIEVI"
123252..123527
/gene="SAV1112"
CDS
123252..123527
/gene="SAV1112"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57274.1"
/db_xref="GI:14246882"
/translation="MAKQATMKNAALKQLTQDADEILHLIKVOLDNLTLPLSCPLYEEV
LDTQMGLQKVEDFAVKLGLVDRECKQIMLRLEKLSKLHEATLV"
123946..125067
/gene="SAV1113"
CDS
123946..125067
/gene="SAV1113"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57275.1"
/db_xref="GI:14246883"
/translation="MVYSAMVPATKGTLTGGIDVPGTFFYNRQLAYIVMSFTIVFFI
AFLMNVKLLSNIKVKQGMITIVSVLLLLLVIGKIDNGSKSWINLCFPMLOASELLKI
AIIYIPFMISKMPRLVSKPKLILSPIVLALGCTFLFLQKDVGGOTLILITLVAII
FYSIGVKNVLRGIPAVLGLFWFVIALMAGWLPSYLARSTLDDPQFESGCTGYH
INSLLAINGGVGFGKLGNSAMKGLYLPPEPTDFIFAICEELGIGGLLVLTLEFF
IVYRFAQFANTSSYFYKLVCGVIATYFGSOTFVNIGGISATIPLTGVPLPFISFGGS
SMISLSIANGLLIIVGKIQVDQQRKQOQKQVDIRQPN"
125621..129073
/gene="pycA"
CDS
125621..129073
/gene="pycA"
/notes="SAV1114"
/codon_start=1
/transl_table=11
/product="pyruvate carboxylase"
/protein_id="BAB57276.1"
/db_xref="GI:14246884"
/translation="MKQIKKLLVANRGEIAIRIFRAAAELDISTVAIYSNEDKSLIHR
```

YKADESYLVGSDLGPAESYLNIERIIDVAKOANVDIAHPGYGFLSENOFARRCAEEG
IKFTGPHLEHDMFGDKVKARTTAIKADLPVPGTDGPIKSYELAKEFAEEAGPLMI
KATSGSGKGMRIIVRESELEADAPRAKSAEAKSFGNSEYVIERIINDPNHIEVQVIG
DEHGNIYHLPERDGSVORRQKVHVAPSVGLSPTLRQICDAAIQIMENIKYVNAQT
VEFLVSGDEFFIIEVNRQVQVHTITEMVTDIDIVKTOILVAAAGDLFGEIENPOOK
DITTLGVAICRIITEDPLNDFDPTGTIIAYRSSGGFVRLDAGDGFQGAETSPYVD
SLVLSHTAISFKQAERKVRSLREMRIRGVKTINIPFLINVMKNKFTSGDYTKFI
EETPELLDIOPSLDGRGKTLEYIGNVTINGPNVKEKPKDYLASTPTVSSSKIASF
SGTKLDVGVGPKGVAEWKKODDVLITDITFRAHQSLLATRVTRTKDMINASKTAD
VFKDGSLEMMGGATFDVAINFLKENPWERLERKAI PNVLQMLLRASNAVGYKNY
PDNVIHFVQESKAGIDVFRIFDLSNMDKQVMAVEAGKISEGTTICYTGDILN
PERSNIYLYYKLAKELEREGFHILAIKDMAGLLPKAAAYELIGELKSAVDLPJHL
THDTSNGLLTYKQADAGVDIIDTAVASMSGLTSOPSANSLYALNGPFRHLRTDI
EGMESLHWSVTYTYDFESDIKSPNTEIYOHMPGGQYSNLSQAOKSLGIGERD
FKXDMVRVNFELGDIVKVTVPSSKVVGDMALYVQNDLDEQSVITDGYKLDFPESVVS
FFKGEICQVPNGVKNQLOAVILKGQEALTARPEYLEPDEFKVRLELLEEQCPVTE
QDILSYVLPKVEYIOTRNOYGNLSLDTPTFFFCMRNGETVEIETIDKRLIILK
ETISEPDENGRNTIYAMNGQARRIYIKDENVHTNANVKPKADKSNPSHIGAOWPSV
TEVKVSGVETVKANQPLLITEAMKMETTIQAPFDGVIKQVTVNNGDPTIATGDLLIETE
KATD"
complement(129213. .130124)
/gene="SAV1115"
/complement(129213. .130124)
/gene="SAV1115"
/note="similar to heme synthase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57277.1"
/db_xref="GI:14246885"
/translation="MFCNNKLKWLGVVATLMMTFVOLGGALVTKTGSADCGSGSWPLC
HGALIPFPDPIDTIELSHRAVSALSILMLVLWITAWKHIGYIKELKPLSIISVGLF
LQALLGAARVINOQNDYVLAHFGSISFSVFLTIIFSIDOKYEADELYIKRP
LRRLWMLAIITGYGVTGALVRHADSLAAYGCMPLFPFDLVPHSQDQWQVLFTRIMA
FVFTIMTYIHAVKNYPNNRTVHYGYTAAFLTILVQITGALSIMTNVNLIIALFH
ALFITVFGMTTYFIMLRLSRSDQ"
130561. .131475
/gene="ctaB"
130561. .131475
/gene="ctaB"
/note="SAV1116
assembly factor"
/codon_start=1
/transl_table=11
/product="cytochrome caa3 oxidase homologue"
/protein_id="BAB57278.1"
/db_xref="GI:14246886"
/translation="MMSKEHTLSNISRVNFKELQOIIMKGLVOGNLIPAFAGAWLAV
VMTNHSFLSSIPQIILLMFLGSLTIMGACALANNYXDIDIRIMPSKQNRPTVNNRITD
QNLLLSFGMLVGEICLFLNIPSGVLGMGIVGVSYISWSKRHTTWNTVIGSFP
GAVPPLIGVNAIEGQISLTAIALFLVFCWQPIHFYALAIKRKDEYALANIPMLPSVK
GKRTVSMFNIWLIILLPVLPLINIGVVFVVLATLLNLGIALGLTTFKKNSDQTKW
ATQMETVSLNLYVIFVFLAVIVSLTLI"
131500. .131961
/gene="SAV1117"
131500. .131961
/gene="SAV1117"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57279.1"
/db_xref="GI:14246887"
/translation="MGPVILPITSTICIVISALIAIGRWRLIWKREINKHKNVMAAA
VFALFTFLYASRTIFIGNAFTGGPASIKKYYTIFLFFHINLATIGGILGLVQIITAF
KQKNVHRFGFPFASVIMFTAITGVAVYLLVLYVPGGETTSLIKATGCH"
132288. .133325
/gene="SAV1118"
132288. .133325
/gene="SAV1118"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57280.1"

/db_xref="GI:14246888"
/translation="MKKLIRRVGVFLVGLFIYLFYSPRLKFDVLENPNKGNKVNRS
EQVNNHNHAKENPKPEGVGTGWDKIKVLTSKFGQADRVYPRFGKYVFKDKNSY
YIYVSTKREIYVYATGKVNVSPLKIQHSAEFTNHTSINPEPSFKVDGKKYEFELS
DEDLKTOTLKIYGDIIYAQVYSDQSKKVLVSRFLTKEMLADIBEPYQLNNSNSTSEHNK
RPVQENPOLISLYEVNEMRKLGKLPKINSDLAHTASNNLYEATNNGSDSVEFTE
DALRGOLDKNHVTYKTTAQNQVGFAPNDVPTLIHWSMSDIHRSRLLSKSKYDEMGGDVN
RDYISLIFLEK"
133341. .133508
/gene="SAV1119"
133341. .133508
/gene="SAV1119"
/note="Vaca"
/codon_start=1
/transl_table=11
/product="putative vacuolating cytotoxin paralog"
/protein_id="BAB57281.1"
/db_xref="GI:14246889"
/translation="MINEASLAILDDIDELADMIVASDIYASFQAKQALENNDEHIY
FINHFKIKREI"
133402. .133773
/gene="SAV1120"
133402. .133773
/gene="SAV1120"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57282.1"
/db_xref="GI:14246890"
/translation="MHQIEMHLSKLNRRSKTWTNTFTLSILKSKKYEVMRFQKY
HPDYKVMLETROKRAYEMLDVVMHYKAREMALQHLIDEVVTIKIATAVSEHKVETG
NPFQTSHGSCATGKGCNSCL"
complement(133836. .134762)
/gene="SAV1121"
complement(133836. .134762)
/gene="SAV1121"
/note="similar to glycerophosphoryl diester
phosphodiesterase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57283.1"
/db_xref="GI:14246891"
/translation="MKKVNRRNLNAVLVGAGFLSSLIMKNKVTIKKOTIPAFFKGN
APYIFAHGGMALRPEQTQAFDYAKQLGVDFETDVRILTQDQLIVFHDTVDRITN
GSGKVSAGHTLAELKLDAAHYFKDINGLTPYRGHAHTTILTFDELLKOYPDVYINVDL
KDAPEVSGISAPQIMFDTAENQAFDRVLVTSFYKEQIVRENKIQAGSVAIGASQOE
VTEAFKYLHLLGGRYVQPLAQTFQMFTHFGIDLTSSRFIKWLNDMNLIPGYTGVNSI
DLMDLIYQGAHTIVTDPRDLAQQFKOTIPNK"
135001. .135255
/gene="SAV1122"
135001. .135255
/gene="SAV1122"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57284.1"
/db_xref="GI:14246892"
/translation="MNLIPRTSIVVYLKHKHHERQIRKYGHIHVSNRDRKFVIMYVNE
QVDQIVHKLMLQKYVRHIDGSPYKYLKTKTYEKHEIYN"
complement(135258. .135647)
/gene="SAV1123"
complement(135258. .135647)
/gene="SAV1123"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57285.1"
/db_xref="GI:14246893"
/translation="MKOTFITLGEGLTDLFEFMTIYVNHORIDKTIYFHSPOAENKK
SSVAIIMNPTGNHQAQVYIMINAIKYPYPSDKKFKQMINDCAEKFDILIGIDVQPP
QAFHDLISLYNLIISVLRLOKWIPELQ"
135717. .136259

AKOKKHPKVLILMGVPGSYLVATDKSYIGDLVKITAGSENVIKVKDRQYVSSNTENLNLN
INPDILITLGHGMEFEVAKKMEQKEFKQNDIWKHFRAKVNHNHYDLEEVFPFGITANVDA
DKAMTQLYDLFYKDKK"
/gene= "SAV1134"
145227. .146048
/gene= "SAV1134"
145227. .146048
/gene= "SAV1134"
/note= "similar to ferrichrome ABC transporter"
/codon_start=1
/transl_table=11
/product= "hypothetical protein"
/protein_id= "BAB57296.1"
/db_xref= "GI:14246904"
/translation= "MDSIIDLRPLRILALMVGAMLAVSGALLQAALQNPLAEANIIG
VSSGALMRALCMLEIFQLYFYLLPSFGLGILPFLIIILHRSFRFNNAVSMILVGVA
LFVLNGLLEITLQDNPLMKIPQGLTMKWSDVITLAVSALGLLITLLSPKLNLLNL
DDIARSGFNIDRYRWLTGLLAVELASATVAIVGQALFGLIIVPHVVRKILVGSYRV
LIPSTVIGAWLLIYVADLLGRVIQPPLEIPANAILMIVGGPMLIYLICQSORNRI"
146110. .146844
/gene= "SAV1135"
146110. .146844
/gene= "SAV1135"
/codon_start=1
/transl_table=11
/product= "conserved hypothetical protein"
/protein_id= "BAB57297.1"
/db_xref= "GI:14246905"
/translation= "MRMKRFLTIYOILLVLIIFGKIVQTYIEDKOERANYEKLOQ
KFOMLSKHOEVRPOFSELEKINKDIVGWIKLSGTSLNPVLQGTNHDYLNLDQFER
EHRKGSIFMDFRNLKLNHNLTILYGHVGDNTMFVDLEDYLKQSYEKHKHIEFDN
KYGIQLOVESAYKTKTKDNYIRTDFENDQYQOQFLDETRKKSVINSDVNVTKDKIM
TLSTCEDAYSETTKRVIYVAKIIVKS"
146872. .147186
/gene= "SAV1136"
146872. .147186
/gene= "SAV1136"
/codon_start=1
/transl_table=11
/product= "conserved hypothetical protein"
/protein_id= "BAB57298.1"
/db_xref= "GI:14246906"
/translation= "MAENRLTILTKGTAKDILIERFYTRHGTIETLEGFGFMFTQLEGE
DFDEVKILTVWKSQAFTDLKSDVFAAKHKVRSKNEDESSPIINNKVITYDIGYSY
MK"
147570. .148310
/gene= "SAV1137"
147570. .148310
/gene= "SAV1137"
/note= "similar to rRNA methylase"
/codon_start=1
/transl_table=11
/product= "hypothetical protein"
/protein_id= "BAB57299.1"
/db_xref= "GI:14246907"
/translation= "MEQITSAQNNRKQANKLKKRKRRTKGLALIEGVHLIEBAVQS
GITVITQALFAPRLDQIIIAQOEVEFINMKVAESLSGTVTPQGFAITEKPHYDIS
KAOQCLLDIVDPGNGLGTLIRTADAGMDAVIMEKGTTPDQDKVLRASQGSVFFHP
VMTQDLDTFTQPGNVPYGTALENAVAYKEVTSDDSFPALLNGEGVNPELLAHTTQ
NLIIPIYGKAESLNVAIAGSILLYHLKG"
148690. .149748
/gene= "phea"
148690. .149748
/gene= "phea"
/note= "SAV1138"
/codon_start=1
/transl_table=11
/product= "Phe-tRNA synthetase alpha chain"
/protein_id= "BAB57300.1"
/db_xref= "GI:14246908"
/translation= "MSEQQTMSLKKQALVDINEANDERALQEVKVKYLGKKGVSGL
MKLMKDLNPKRPAFGKVNELRQTQNELDERQOQLVKEKLNKLAEEETIDVSLPCR
HIEIGSKHPHITRTTIEIEDLFLGLGYEVINGVEQDHYNFEMNLNPKSPHARDMODS
FYITDEILLTHTSPQARTMESRHGGPVKILICPKGVYRRSDDDATHSHQFTQIESL

VVDKNVMSDLKGTLELLAKLFGADREIRLRPSYPFPTPEPSVEVDVSCFKCKGKGCN
VCKHTGTEILGAGVHPNVLMEAGFSDSEYSFGAFMGMDRIAMLKYGIEDIRHYFT
NDVRFLDQFKRAVEDRGM"
149748. .152150
/gene= "phea"
149748. .152150
/gene= "phea"
/note= "SAV1139"
/codon_start=1
/transl_table=11
/product= "Phe-tRNA synthetase beta chain"
/protein_id= "BAB57301.1"
/db_xref= "GI:14246909"
/translation= "MLISNWKKEYVTIDSVSDLAERITRTGTIEVDLDIDYTKDIKN
LVGFKVSKKEHPADKLVNQCQVIDGEDEPVQIVCGAPNVDAGOVIVAKVGGRLPGG
IKIKRAKLRSESEGMICSLQEIGISSNYIPKSFESGIYVSESOVQCTDALQALYLD
DQWFEFOLTPNRADALSMIGTAYEVALYNTKTKPETTSNELSELSEANDELTUTVIENE
DKVPYSARVVDVHTIIEPSPIQMAQRLIKAGIRPINNVVDISNVTILLEYGQPLHMFQD
DAISQOIVVRQANEKGKMTLDDTERELTSDIVITNGQTPPIALAGVMGGDFSEVKE
QTSNIVIEGALFDPVSIIRHTSRRLNRSSESSRFEKGIATFEVDEAVDRACYLLQTYA
NGKVKDRVSGSELGAFITPIDITADKINRTIGFDLSONDITVTFNOLGFOFTEINDDV
ITVLVPSRRKDIITKEDLIEVARIVGYDDIPSTLPVFDKVTSGQLTDROYKTRWKE
VLEGAGLDQALITYSLVSKEDATAESMOORQOTIDLLMPSEAHASLRQSLPHLIEAAS
YNARKNKDKLFEIGNVFFANGEGELPDQVEYLSGILTGDYVVMQGGKRETVDFFYL
AKGVDRVSEKLNLEFSYRADIDGLHPGRTEAILLENKVVGFIEGLHPTLAAANDLK
RTYVFEINFDALMAVSQYINQPIPRFPGMSRDIALEVDQNIIPAADLLSTIHAHGN
ILKDTLVFDVVQGGHLEKKGKKSIAIRLNYLDTEETLTDERVSKVQAEIAEALIBQAV
IR"
complement(152582. .153520)
/gene= "SAV1140"
complement(152582. .153520)
/gene= "SAV1140"
/note= "similar to ribonuclease HIII"
/codon_start=1
/transl_table=11
/product= "hypothetical protein"
/protein_id= "BAB57302.1"
/db_xref= "GI:14246910"
/translation= "MANIVEKLSDKDITTLMSRTTFDTENLPQGMKARAKYONTVTVNI
YQSKVMFGNHAFVASKELLPOHSOLNTNKKKKKNMANSLEOTLMYDOFNCIGSDE
AGSDYGPGLTVCACAFVTKHEVPILKTLGVDDSKKLTDTDKTIVELAEQALVAFIPHSLLT
LHNDKYNQAKGTQVMKRAVLHNEAIKNVLEKIDSQOLDYIVIDQFAKREVYSHYA
LSDIPLPKTKYFKETKESKSLATAVASIIIRYAEITTYMDQISQISKYINMTIPKGAGAKVD
VIAAKIIKKYGLSLRDLTISKKHFKNREKAQKILKPL"
153890. .154156
/gene= "SAV1141"
153890. .154156
/gene= "SAV1141"
/codon_start=1
/transl_table=11
/product= "conserved hypothetical protein"
/protein_id= "BAB57303.1"
/db_xref= "GI:14246911"
/translation= "MAQFKKNVVSINDQLFTIVGEDNPHEIRYVAHLVDDKIKELGY
KAGLDTSRKAILTAVNIMHEKVLLEENRRLKQOIHKKQOREQ"
154157. .154678
/gene= "SAV1142"
154157. .154678
/gene= "SAV1142"
/codon_start=1
/transl_table=11
/product= "conserved hypothetical protein"
/protein_id= "BAB57304.1"
/db_xref= "GI:14246912"
/translation= "MVIDFIIIIFFVYVIVGPRGFWLSMHLISATIVSIWIASQFY
KSIVERLIVPIYPKPTAFNTTAFENHLQNRFEAIVAFMLTFLCKFLYLIIVTF
DKIIAYONIHIFSRAMGMVGVFMTIIVLHFTLYLLALYPNEALQHOLKISIVSHSLI
FHPYLSAFTNL"
154751. .156463
/gene= "SAV1143"
154751. .156463
/gene= "SAV1143"
/codon_start=1

/transl_table=11
/product="DNA-dependent DNA polymerase beta chain"
/protein_id="BAB57305.1"
/db_xref="GI:14246913"
/translation="MTKKDVIKLLEQIAYMYELKMGENTFKISAYRKAASLELDERPL
DEISDYTELKGIGKGAEVINDYRETGESQYLQOLQEEVPEGLIPLKIOGLGSKKIA
KLYKEINLVKASQVACENGKVSSELSGFAKTEQNIIEAVKQLGAKKORYPIDQMR
LNOEIIDYDITLNYIDQYSAGFRFKEMSKDLDFIISTDPKAVQQOQLLNPKNVK
EVAVGNTKYSLELAYDDETIGVDFRIEPRAFYHTLQHTFTGSKHEINRIRQLAKARDE
KVSYEIGEQADGTLIOYDSEAKIYEHFNVNFIIPAMREDGSEFDKDLNITLDDING
DIHMHTYSDGAFSIRDOMEVANIACYKGMVITDHSQSLRVANGLOVERLLRQNEIK
ALDKSESDIYSGTEMDLIDPGSLDYDDEILAQLDYVIGAIHQSPNQSEQIMERIA
NACRNPYVRHIAHTGRIIGRRDGYKPNIEQLMALAEENTVLEINAPHRLDLNAEI
VRKYPNVKUTINTDAHTHTNHLDFMNYGVATAQKGFVTKDRVINALSREAFKDPFIENNI
KLKK"
156473..158821
/gene="mutS2"
156473..158821
/gene="mutS2"
/note="mutS2"
/note="SAV1144"
/codon_start=1
/transl_table=11
/product="MutS-like protein"
/protein_id="BAB57306.1"
/db_xref="GI:14246914"
/translation="MRQKTLIDLVEFEKIKSLVANETISDLGLEKVNQMPPATNFETVV
FQMEETDEIAQIYNKHLRPSLSGLSKVSAP IHRADIGGVNLVSELNLIKRLIQVQNF
KTYNOLRVEDEGVKYPIILDDKNOLPVLTDLPHQINETCDTDVLDYNASEYELQGIS
KISSTNORLONDRILVKSOAKOKLSDAIVTVYRNERNVI PVKAEYRQDPNGVILHQOS
ASQOTLIERPSVVENNQISLRDHEATEKERVLTQLTGVAAKDALLVAEQVMGQ
LDPLIARAKYSRIKGTGPIFKEERTVYLPKAYHPLLNRETVNIETFEMEDIETVII
TGPTNGKVTYKTLGLIIVMAQSGLLIPTLDGSLSVFNKVCYDGDQESIQESLST
FSPHMTNIVEILKHADKHSVLVDFELGAGTDPSGAALAMSILDHVRKIGLSVMAITH
YPLKAYSYNREGVMNASVEFDVDTLSPYTKLLMGVPGRSNAFDISKKLGSLNAINLK
AKTMICTDSEKINEMIESLERNVKRYETORLEDRDLVKSAEVOVHDLKSQYQOFQNYE
KSLIEAKKAKOKIKAAATKEADDIILKQLREQGADVKEHELIDKKRLDODHTEA
KSTQNQKQDKYIVAGDEVKLVSQKGEVLEIVNDEEAI VOMGIIMKKLPDIEDLE
KKQEKVKPTKMTVRQNRTIKTELDLGRYREDALIELDQYLDQAVLSNYEQVYIIH
KGKTALQGVQOHLKKHKHSVSDFRGMPSEGGFVTVATLK"
158994..159308
/gene="trxA"
158994..159308
/gene="trxA"
/note="SAV1145"
/codon_start=1
/transl_table=11
/product="thioredoxin"
/protein_id="BAB57307.1"
/db_xref="GI:14246915"
/translation="MAIVKVTADDFDSKVESGVQLVDFWATWCGPCKMIAPVLEEELAA
DYEGKADILKLDVDENPSTAAKYEVMSIPTLIIVFKDGPVDKVVGFQPKENLAEVLDK
HL"
159632..161413
/gene="uvrC"
159632..161413
/gene="uvrC"
/note="SAV1146"
/codon_start=1
/transl_table=11
/product="excinuclease ABC subunit C"
/protein_id="BAB57308.1"
/db_xref="GI:14246916"
/translation="MEDYKQRIKNKLVNVPMEPGCYLMDRNDQVLYVGKAKLRNRL
RSYTCGAHDAKTRLVGEIRPERFEIVTSETESLLELNLIKOYQRYNILLKDDKSY
PFKIKTEKYPRLVTRVQGTGYFGPPYNAISAQETAKLLDRILPYPKCDKMPDP
LCLYIHGCLGCPYVDVLSKYAQMTKEITDFLNGEDKTIILKSLERMLTASESLDF
ERAKERYDITQIONTRKQIMSSDKTIDFVGYCVKDWCMCIQVFTIRQGNMIRKD
TTPMIQQTBEFEETFFIGQFVSLNQHILPKVEHVPNRNLDKEMIQSVDTKIVQPARG
PKKDMDLAAHNAKVSNNKFELISRDESKT IKAIEELGTQMGCIOTPIRTAEAFDSNI
QGVDPVSAWTVFDGKPKDKNYRKYIKTVKGPDDYKSMREYRRYSRVNLNEGRLPP
DLIIVDGGKGHMGVLDVLQNELGLDIPVAGLOKNDKHQTSellyGASAEIVPLKNS
QAFYLLHRTODEVHREAITFTROTQTKTGLKSLTDIDIGGNKRRTLLLRSGFSIKKM
KEATLDFDKNIGIPENVAKNLHQHLK"

gene
CDS
161737..162351
/gene="sdhC"
161737..162351
/gene="sdhC"
/note="SAV1147"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase cytochrome b-558"
/protein_id="BAB57309.1"
/db_xref="GI:14246917"
/translation="MAQSKNEFYLRRLHSLGLLPIGAFVLVHLLVNHQATQCAEAFN
KASNFMESFLIIVFLFYIIPLLYHGLFGIHIAFTAKENGVGHYSIFRNMFFHQFQ
SGLITFTIGTGHLOWRLQKAFYKQEVNMDLMHETLOHPGMAIFYIICIIIAVVFHFQF
GLMSFLVTWGLQSPKRSQRFVTVWSLIVFLVSIYIGVTAIIAFM"
162403..164169
/gene="sdhA"
162403..164169
/gene="sdhA"
/note="SAV1148"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase flavoprotein subunit"
/protein_id="BAB57310.1"
/db_xref="GI:14246918"
/translation="MAEKHLIVVGGGLAGLMSTIKAAEKAHVLDLVSVVPVKRSHVC
AQQINGAVNTKGEQSPWIHFDDTVYGGDFLANQPPVKAMTEAAPKIIHLLDRMGVM
FNRTNGLDIDFRFEGGTLHRTAYAGATTQQLLYALDEQVRAVEVDGLVTKYEGWEF
LGIVKDDDSARGIIVAQNMHTAETETPGSDAVIMATGPGGI IFKKTNTSNLWTSASA
IYVQQAIIYANGEFIQIHPTAIPGDDKRLMLSESARGEGRINTYTKDGKWPYLEEKY
PDTGNLVPRIATREIFDVCINQKINGENMVYLDLSHDKDELHVLQGLGIIETEKY
FTGDDPRKVPKIPFAPVHYVSMGLLYVDYDQMTNIKGLFAAGECDFSQHGNRILGANS
LSAIYGGTVAQNAIDYISNDRSYTDMDESIFEKRAEQEEDFKLLAOMRGTENAYK
LHRELGIAMTANVTVVRENEKLLLETKKIVELMKRYEDIDMEDTQVWNOAVFTFROL
WNMLVARMTTIGAYNRNESRGAHYKPEFFERNDENLWKTMTMASFOCAFEKPKQTYIDD
VDVSLPPRKEDYTSKSGGKK"
164184..164983
/gene="sdhB"
164184..164489
/gene="sdhB"
/note="SAV1149"
/codon_start=1
/transl_table=11
/product="iron-sulphur subunit of succinate dehydrogenase"
/protein_id="BAB57311.1"
/db_xref="GI:14246919"
/translation="MKNTPOHETQSKPKQTKVLIIKRODTSKPKVEETFEIPLYRNM
LVNIACLMIEIRNPVNKIGKTTPVVWDMNCLEVCACSWINGRARQSCRLRLIN"
164393..164983
/gene="sdhB"
/note="SAV1150"
/codon_start=1
/transl_table=11
/product="iron-sulphur subunit of succinate dehydrogenase"
/protein_id="BAB57312.1"
/db_xref="GI:14246920"
/translation="MGYELLRRSMWSMFYGYQWSCKTILSAIVDLQBPQIRLEPMTNFF
LVICDLQVDRSRMFDNLKRMKAWIPIDGTYDLGPPRMEPKKQOTAYELSKMTCGVC
PEVCPNVTENKFPVGAQISQVRLFNLRHPTGSKTKOERLNALMGTGGLQCGNSQNCV
NACPKGIPLRTISAAMNRETTTFHMFKSFPGSDHEVE"
165219..166019
/gene="murI"
165219..166019
/gene="murI"
/note="SAV1151"
/codon_start=1
/transl_table=11
/product="glutamate racemase"
/protein_id="BAB57313.1"
/db_xref="GI:14246921"
/translation="MKNKPIGVDSGVGLTVAKMEIRQLPNETIYLDIGRCPCYGR
PGSQKQYVTEIARKLMEFDIKMLVIACNTATATAVLEYLQKTLISIPVIGVIEPGARTA
IMTRNQNLVLVTEGTIKSEAYRTHKIRNPHVEVHGACPGCFVPLEQMSYSDPTI
TSIVIHQTLKRWNRNSEDVTILGCTHYPLLYKPIYDYFGGKKTIVISSGLETAREVSAAL

gene

166031..166618
/gene="SAV1152"
CDS
166031..166618
/gene="SAV1152"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57314.1"
/db_xref="GI:1424692"

/translation="MKREIVASNQGKINDFKVIFPDYHVIGISELIPDFDVEETGST
FEENAILKSAAKALKNTVIADDSGLEVALNGEPGIYSARYAGENSDEANIERKLL
NKLNTITDRAQFCVVISMSGPDMETKVFAGTVSGEJADGJGENGFGYDPIFYVPLK
DKTMAQLSKQKQGOISHRRNAINLLQAFLEGKNV"

gene

166611..167114
/gene="SAV1153"
CDS
166611..167114
/gene="SAV1153"
/codon_start=1
/transl_table=11

/product="conserved hypothetical protein"
/protein_id="BAB57315.1"
/db_xref="GI:1424692"

/translation="MSKWIIVSDNHTESGVLYQIYEMHPDADYVHLGDSFPAYDOTE
LSLFNRVKGCDYFEPENAVAKINDVKAFYTHGHLYQVNRTRDLLAEGLEGCLF
AFYGHTHVAKYEYINGVHVINPGSISQSRSSMEETAEVIIDQTLHGTFNFKNRHHE
TISHTTF"

gene

167272..167409
/gene="SAV1154"
CDS
167272..167409
/gene="SAV1154"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57316.1"
/db_xref="GI:1424692"

/translation="MVDNPIWLNVHDVLYKSVLLKSIKLRFEQVRKFEILVKIMNRH
I"

gene

167603..167932
/gene="SAV1155"
CDS
167603..167932
/gene="SAV1155"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57317.1"
/db_xref="GI:1424692"

/translation="MKNFIGKISILSIAISLTVSTFAGESHAQTKNVEAAKKYDQYQ
TNFKKQVNRKVVDQAQAVNLFRKRTVATHRAQRAVNLHFQHSYEKKKLQRLQIDL
LKYNILK"

gene

complement(168301..168702)
/gene="SAV1156"
CDS
complement(168301..168702)
/gene="SAV1156"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57318.1"
/db_xref="GI:1424692"

/translation="MKNKTKTIIASTVTAAGLLTQNDKAKAFESYEWKGLETAKNLA
DQAKKDERIDLKESDKNLTPYKAEITYNDLVIYKLSQGDVKKAVVRKIDGGPRD
YTFDLTRPLEENRNKNIKVVKNGEIDSYYWD"

gene

169372..169878
/gene="SAV1157"
CDS
169372..169878
/gene="SAV1157"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57319.1"
/db_xref="GI:1424692"

/translation="MIKILITILLCILSYLVNRRKKPFLFKTLFMGVVFIFIGYI

gene

170138..170635
/gene="SAV1158"
CDS
170138..170635
/gene="SAV1158"
/note="similar to fibrinogen-binding protein"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57320.1"
/db_xref="GI:1424692"

/translation="MKNKLIASLLTIAAIGITTTTIASTADASEGYGPREKKPVSIN
HNIVEYNDGTFKYOSRPFNSTPKYIKPKHDYNILEFNDGTFEYGARPOFNKPAKTD
ATIKKEQLIQANLVREFEKTHTVSAHRAQKAVNLVSFEYKVKVQLQERIDNVLK
OGLVR"

gene

170785..171135
/gene="SAV1159"
CDS
170785..171135
/gene="SAV1159"
/note="similar to fibrinogen-binding protein"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57321.1"
/db_xref="GI:1424692"

/translation="MKPKVILGTALLSSGTIATIEGNKADASSLDKYLTESOPH
DKRIAEELRTLLNKSNNYALAGSLNPYYKRTIMMEYRAKALKKNDVSMADAKVA
LEKIYEIDEIINR"

gene

complement(171901..172086)
/gene="SAV1160"
CDS
complement(171901..172086)
/gene="SAV1160"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57322.1"
/db_xref="GI:1424693"

/translation="MKHINKKATKLEEDYKKVQQRSEETDRGCLKSVIMLIVLFFL
IIFGIACSTNNVFFQ"

gene

complement(172708..172941)
/gene="SAV1161"
CDS
complement(172708..172941)
/gene="SAV1161"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="BAB57323.1"
/db_xref="GI:1424693"

/translation="MVFNDWEINWVYSGEHVVTNEKSLFVIDEYDAIAIYYLDGK
LKVAHVNYGSDFTIDVANKVFELNIQNPNDEH"

gene

173460..174779
/gene="tnp"
CDS
173460..174779
/gene="tnp"
/note="SAV1162"
/codon_start=1
/transl_table=11

/product="transposase for insertion sequence IS1181"
/protein_id="BAB57324.1"
/db_xref="GI:1424693"

/translation="MCNDTLELRKIDENIKYINQIEDVIIKGGKATVVNAVLTYPKS
ACYCCGVKNQGOJHKHGKRVSRITLLKTOGYNTYLNLAKEKFCLENGTFTAKTSIV
DESCFISRCVTKQVIEATKVKTEIDTAEDNCISPTSVSRIRTKAANSLRIKPFNCPL
EHIADEDFKSVKNVTGMSFIFIDNTHDIVIDLENKTRTFLRAYFERFDLKNRQVQK
TVTIDMEYVRLFRDLFPNAAIFDRFHIVQHLNRELKNKYRVQVWNEYRNKKGPNTY
IFKNKSVLIMDTSKTIFSKRWKSKFAYKRSDDIVEFMSLDDILRRSYELVQGLR
KDLRLCNWPKFINRLNSVKKSVKGVKVKYKQHRMLRNTIYYPAFNNGAIEGI
NNKIKLIRISFGYRNPNFKARIMTFSLYKGEKKKTKTPNNGLAA"

gene

complement(174938..175897)
/gene="SAV1163"
CDS
complement(174938..175897)

/gene="SAV1163"
/codon_start=1
/transl_table=11
/product="alpha-hemolysin precursor"
/protein_id="BAB57325.1"
/db_xref="GI:14246933"
/translation="MKTRIVSSVTMTLLLSILMNPVANAADSDINKTKTTDIGSNT
TVTKGDLVTYDKENGHHKVFYSFIDDKHNKKLLVIRTKGTAGQYRVYSEEGANKS
GLWPSAFKVLQDPDNEVAQISDYIPRNSIDTKYMSLTITYGNGNVTGDDTKGKIG
LIGANYSIGTLKYVQDPDTILESPTDKVKVKVIFNNMVNONWGPYDRDSWNPYVG
NQLFMKTRNGSMKAAENFLDPNKASSLLSGFSPDFATVITMDRKAASKQQTNIDVIYE
RVDDYQLHWTSTNWKGTNTKDKWDRSSERYKIDWEKEWTN"
176747..176944
/gene="SAV1164"
176747..176944
/gene="SAV1164"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57326.1"
/db_xref="GI:14246934"
/translation="MTELNNIINSLQSLFESESGYKISKNSGVPYQTVQDLRNGKTKL
EDARFRTIKLYSYVYSLKEH"
complement(177171..177302)
/gene="SAV1165"
complement(177171..177302)
/gene="SAV1165"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57327.1"
/db_xref="GI:14246935"
/translation="MTEMVFKIYFVVTQQALSVEFLFEFLXVGAPPTCIVCRISP"
complement(177434..178150)
/gene="SAV1166"
complement(177434..178150)
/gene="SAV1166"
/note="similar to exotoxin 1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57328.1"
/db_xref="GI:14246936"
/translation="MSKNITKNIILTTLLLTGTVLPQNKVPFSFYSEAKAYSIGOD
ETINELIKYITQPHFSFNKWLQYDNGNIYVELKRYSSAHTSLWGAESWGNINOL
KRDYDVFGLKDDTDQLWMSYRETFGGVTPRAKPSDKTYNLFVQYKDKLQRIIGH
KIYQGNKPVLTLEIDFRAREALIKNKILYTNRNKGLKITGGNNITIDLSKRLHS
DLANVYVKNPKITVDVLFDP"
complement(178258..178983)
/gene="SAV1167"
complement(178258..178983)
/gene="SAV1167"
/note="similar to exotoxin 4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57329.1"
/db_xref="GI:14246937"
/translation="MNNITKKIILSTLLLTGTAFTQFPNTPINSSSEKAKAYINON
ETWNELTKYSQKYLTFSSNLQKDNGTIHATSLQFSWYSHQVYGPESWGNINOL
RNSVDIFGKIDQETIDSFALSOETFTGGVTPAATSNDRHKYKLNVTYKDKAETFTGGF
PVYEGNKPVLTKELDPRIROTLIKSKKLYNNSYNKGIKITGTDNNYITIDLSKRLPS
TDANRVVKKPQNAKIEVILEKSN"
complement(179078..179803)
/gene="SAV1168"
complement(179078..179803)
/gene="SAV1168"
/note="similar to exotoxin 3"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57330.1"
/db_xref="GI:14246938"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

/translation="MKNINMKLVLTALLLLGTTSTQLPKPTISFSEAKAYNISEN
ETINELIKYITQPHFSLSGKWLKQPNKSGSIHATLOTWVWYSHQVFSWGNINOL
RNKYVDIFGTKDETDVEGYTYDETFTGGVTPAATSSDKPYRLFLKYSDKOQTIGGH
EFYKGNKPVLTLEKLDPRIOTLKNKKLYNGEENKGIKITADGNNYITIDLSKKLKL
TDNRVYVKNPKNAKIEVILEKSN"
180241..181242
/gene="argf"
180241..181242
/note="argf"
/gene="SAV1169"
/codon_start=1
/transl_table=11
/product="ornithine carbamoyltransferase"
/protein_id="BAB57331.1"
/db_xref="GI:14246939"
/translation="MKNLRNRSFLTLTLLDFSROEVEFLLTJLSEDLKRAKYIGTERPMLK
NKNIALLEFKDSTETRCAFEAHDOGANVTYLGPTSQMGKKETTKDTARVLGMYD
GIEYRGFSORTVETLAETSGVPVWNGLTDEDPHTQVLADFLTAKEVLKKDAVLFNIFY
VGDGRNNVANALMOGAAMNMFHLVCPKELNPTDELLNCKNTAAENGNGNLTDDI
DQVKGSDVIYTDVWVMGEFDEYMKERLELLKPYQVKNKEMDKTGNPNVIFEHCLPS
FHNADTKIGQIFEKYGITREMEVTDEVFESKASVVFQEAENRMHTIKAVMVAATLGEF"
181265..182197
/gene="SAV1170"
181265..182197
/gene="SAV1170"
/note="similar to carbamate kinase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57332.1"
/db_xref="GI:14246940"
/translation="MAKIVVALGNGNALGKSPQEOLELVKNKTAKSLVGLITKGHEIVIS
HGNGPOVSGINLGLNYAAEHNOGPAFFAECGMSQAYIGYLOESLQNELHSGIMDK
QVWTLTVQVEDENDPAFNPNPSKPIGLFYANKEAEQIQKEKGFIFVEDAGRGYRRVP
SPQISIIELIESIKTLIKNDTLVIAAGCGGIPVIREOHDFGKDAVIDDKDKTSALLG
ANIQCDDLIILTAIDYVIINFNTENQQPLTKTNTVDELKRYIDENQFAGKSMPLKIEAA
ISFIENPNKGSVLITSLNELDAALEKRGVTVIKK"
182369..183925
/gene="SAV1171"
182369..183925
/gene="SAV1171"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57333.1"
/db_xref="GI:14246941"
/translation="MENTINSEKKRFLKMPGAFMLFLTIVVAVIATWVIPAGAY
SKLSEPSQOELKIVPNHNQVKYPGTQOELDKMKVKIKIEQKSGAINKPVSIPNTY
ERLKHQHPAGPEQITSSWVEGTIEADVIMVFTLVGLGLIGVGVQASGESGLLALTKKT
KGHEFLIVFVSILMIIGGTLGIEEBAVAFYPILPFIATGADYDSIVSGAIFLASS
VGSTFSTINPFSVVIASNAAGTFTDGLYWRFIGACIVGAIFVISLYWYCKTIKNDPK
ASKSYEDKAFEQQWSVLKDDSDAHTLTKKILITLFLPFPIMVGMVQGWFPVM
MISDTILHFSSLSVQHMSGPLFIIVLFIFFCLGFIVPSSGLAVSLIIGLARGINLVNEG
GIPRFVITVYQGYAMLFAPTLGVMATLQMLNMRYSWHFREVWVPVAVLIFGGG
VLITQVLIYS"
184232..184459
/gene="SAV1172"
184232..184459
/gene="SAV1172"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57334.1"
/db_xref="GI:14246942"
/translation="MTHLTRVLDLTGICVWLFSKYFVAYANMVFDNLRWYLLLENI
PHLPILFILMFIFGVPSEMIKDRQRKNNGV"
complement(184765..185712)
/gene="SAV1173"
complement(184765..185712)
/gene="SAV1173"
/codon_start=1
/transl_table=11

gene
CDS

gene
CDS

gene
CDS

/product="conserved hypothetical protein"
/protein_id="BAB57335.1"
/db_xref="GI:14246943"
/translation="MRLQKAPLVTSGLVLGLGLGNLLKDLSTLNAVCGTFAFLWI
HLCTMIKYPNNVKQSLNSPLVSVSFTTFMGSFLGTYTLNTFFSNITFTNSLITPIW
ILCLVGIMTHMIIFSILYKLDLSLENVYPSWTLF IGIAIAGTAGTAPVSGCGFFIGOLTV
IYGFAVCIVPLIVFRKLKAPFSLKPNSTICAPFSLVAAAYJAFPRANAFIVI
IFILLAQIFYIILIQPLKLLKPFSPVSAFTFPDVISATALKNSLPLVLMFPDIWKG
LLEFIEVLATVIVLRVFIQYLHFFLKENQDKFLRNASQ"
185961..186149
/gene="SAV1174"
CDS
185961..186149
/gene="SAV1174"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57336.1"
/db_xref="GI:14246944"
/translation="MRNQIQKLLDSDLSSLIHSKQTGVPOSTIHRMKKERSLDNMSL
KNALFYKANSIFSNEH"
complement(186617..186690)
/gene="trna-Arg"
/note="SAVrNA16"
complement(186617..186690)
/gene="trna-Arg"
187512..188198
/gene="SAV1175"
CDS
187512..188198
/gene="SAV1175"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57337.1"
/db_xref="GI:14246945"
/translation="MGYKNLIDFDDTIVDFYDAEWAETHYMANVENHKATKDDLTFF
KINHOHWAFOQNKLTKSEVLSERENVYKHHQMEVDGHRADVLRNGLAELAKYVF
DOTLETIVLSRHDLYIYVTGTEQKRRLNQTPLHKYIKKIFISEETGYQNPPEF
FNIVFNDIGEDERQHSIIYVDSLSLILGGINAGIATCFNFRGFDHNPGLIIPDIYN
SWQLNDIVR"
complement(188302..188748)
/gene="SAV1176"
CDS
complement(188302..188748)
/gene="SAV1176"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57338.1"
/db_xref="GI:14246946"
/translation="MSEIKRLEINYKTDELFFENFRAFGNKKDLYMVNQLNCOMIDASSD
SPFYGIFVGDQLGARNALLKGDVEIYFPDFEDYILLWKLEVLPKPYQNKGYASELID
FAKSPNPIKAIGRNDSDKOFFLHHGFTDVEAKNIEGHDVLLMKPIR"
188943..190556
/gene="SAV1177"
CDS
188943..190556
/gene="SAV1177"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57339.1"
/db_xref="GI:14246947"
/translation="MDCKVVSUNEKDOFTPKIKSSDPVITGLFOYDAAOQTSEKRRMS
KENNGREALANVIREMSDLKSNQEQLNIQHLANGSKVYIGGQOAGLFGGLPYTFH
KIFSITIUSKELTDTHKQVVPVFIAGEDHDFDEYNHTFVYNNENHSLHKKYKHTME
MPEFTVSRYPDKALQPTLKMFIHMKETVHTQGLLEICDRIIDQYDSWIDMFKALL
HETFRAYGLVFDIAQOPEPLRKMFAPEMKILKKHOLLDDAFRATQORTQONOGINAMIQ
TPTNVHLFDENMRQLVSVDGKHFLNKTKTYIKEELIINAENOPELFSNNVYTRP
LMELEKLVTVAVGGPSEIKYWAEUKDVFELEDVEMPVMPRLRITLYLNDRIEKLKLSK
YNIPEKLVDVGGVEGSKAFIREQASHQFIKEKVEGMIEQOORLNKLDLDEAVGNNQNI
NLVNKNNEHIHQYDYLKRYLLNIERENDISMKQPREIQETLHPMGGIQLERINWPLQ
LINDFTDVFKEPSTYPLSYTFDHIHKP"
190700..191131
/gene="SAV1178"
CDS
190700..191131

/gene="SAV1178"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57340.1"
/db_xref="GI:14246948"
/translation="MFMGEYDHLQDTRGMTIPSKFRYDLNERFIITRGLDKCLFGYT
LDEWQOEERKMTLPMTKDKARKFMRMFFSGAVEVDLQKGRITPNQLRKYANLTKE
CTVIGSGNRIEIMDRETWDFEYESERSFEDIAEDLIDFDE"
191147..192082
/gene="SAV1179"
CDS
191147..192082
/gene="SAV1179"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57341.1"
/db_xref="GI:14246949"
/translation="MFHHSYMLNETIDYLVNKENGVIYDCTLGGAGHALYLLNQLND
DGRLIAIDQDTAIDNAKEVLKDLHLKTVFVHSNPRELTQILKDLNIEKVDGIYDGL
VSPQLDIPERGFSYHHDDATLDMRMDQELTAYEIVNNMSYKALVIFYRGEKFS
KOIARRIEAHRQOQPIITTTLELVDI IREGIPAKARKGGHPAKRVQALRIAVNDELS
AFEDSIEQALIELVKVDCGRISVITFHSLEDRLCKQVFOYEYKGPVPRGLPIPEAYTP
KLXRNKRKPITAREEDDDNNRARSAKLRVAEILK"
192096..192497
/gene="ftsL"
CDS
192096..192497
/gene="ftsL"
/note="SAV1180"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="BAB57342.1"
/db_xref="GI:14246950"
/translation="MAVEKYQPYDEOVYNSIPKQOPQTKPEKTVSRKVVVQLTKFE
KVYITILITVAMLSIYMSLSKMDAYDTRKGIADLDIKIDKQSSENSALQSEIKRNS
YERIERAKQIGMSLENDNNVKKVRSNGEAKN"
192478..194712
/gene="pbpA"
CDS
192478..194712
/gene="pbpA"
/note="SAV1181"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein 1"
/protein_id="BAB57343.1"
/db_xref="GI:14246951"
/translation="MAQKIKIKKNKIGAVLLVGLFGLLPILVLRISYIMTCHSNG
QDLVMKANERYLVKNAQPERGIYDRNGKVLAEDEVRYKLVAVIDRKASANSKKPRH
VVDKETAKKLSIVIDMKPEIEKRLSOKKAFQTEFGKGTNLTQYDKLKTERRNLPNG
ISLLPETEREYPNGFNASHLIGRAKNPDTGELKALGALGVEKIFDSYLSGSGSLRYIH
DIWGYIAPNTKKEKPGDDVHLTIDSNIQVFVEEALDGMVERYPQKDLFAVYMDAK
TGEILAYSORPTENPETGKDFGKKWANDLYXONTYEPGCTPKSYGLAAIOEGAFDDPK
KYKSGHRDINGSRISDNWRVWGGEIPMSLGTYSSTNLMHLQDLVGNADKHKSWYERF
GFKSTGMFDGEAPGOIGWSNELQQKTSFGSTTTPVQMLQAQSAFFNDGNMLKP
WFYNSVENPVSKRQFYKGQKQIAGKITTKTAKEVKOLDLWNSKKSHAANYRIDGY
EVECKTGTQAAVAPNGGGYVKGPNPYVFSFGADPKPKNKPIVIVAGKSLAQKNQDEAY
ELGVSRAKFPIMENTLYKLVNKGSKDQTSNAEYSKVPDVGODKQKADINYSKASLEP
VTIGSGTOIKAQSIKAGNKVLPKHSKVLLLTDGDLTMPDMSGTKEDEVIAFENLTNIKV
NLKGSFVSHQSISKQKLETKDKIDVEFSESVDSNSTNNSDSNSDDKKSDSKTOK
DKSD"
195004..195969
/gene="tray"
CDS
195004..195969
/gene="tray"
/note="SAV1182"
/codon_start=1
/transl_table=11
/product="phospho-N-muramic acid-pentapeptide translocase"
/protein_id="BAB57344.1"
/db_xref="GI:14246952"
/translation="MIFVYALLALVITFVLVPLVPLIPLTKRMKFGQSTREGGPQSHMKK
TGPTMGGLTFLLSIVITSYALLVAIFVDQANPIILLLVFTTIGFGLIGFIDDIYIVVKKN

gene
CDS
195971..197320
/gene="murD"
195971..197320
/gene="murD"
/note="SAV1183"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylmuramoylalanine-D-glutamate ligase"
/protein_id="BAB57345.1"
/db_xref="GI:14246953"
/translation="MLNTYKLENKLVVLGSLAKSGYEAALLSKLGANVTVDGKDLS
ODAHKLESMGIVSVSGSHPLTLLDNPNLIIVKNPGIPYTVSIIDEAVKGLKILTEV
ELSYLSEAPLIATVTGNTKGTVTYSLIGDMFKSKRLGRLSGNIGYVASKVAQEVKPT
DYLVELSSQLLQIEIKYKPHIALIITNIGMSYSAHLDYHENLENYQNAKQIYKNOTEEDY
LICNYHORVIESEELKAKTLFYSTQOEVGDIYKDGFIYKGVYRIINTEDLVLPESH
NLENILAAVLACILAGVPIKAIIDSLLTFSGIEHRLQYVGTNRTNKYINDSKATNTLA
TOPALNSFNQPIIWLGGDLRGNEFELIPYMNVRAMVVGOTKAKFAKLGNQSGKS
VIEANVVEDAVDKVQDIIIEPNDVLLSPACASWDQYSTFEERGEKFIERFRAHLPSY"
197336..198655
/gene="divIb"
197336..198655
/gene="divIb"
/note="ftsQ homolog
SAV1184"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="BAB57346.1"
/db_xref="GI:14246954"
/translation="MDKTKNDQOESNEKDLELETRNTSKRRQKRKSKATFHSNQ
NKDTSQOAFDEEYILINKDFKRESNDNNGSSASHANDNDDSDTSNENEDYR
YNEIDDDNESGIAVNGEQPSAPKEQSDSNDDEETVTKERKSKVTQLKPLTLEEK
RKEDRRKQRIQYSVITILVLLIYVIMFSLPSLKIAHVNINGNNHVSTSKINKVLG
VNDNSRMYTESKKNAINLDEEDPLIKSVETHKQLPNTLNVDITENEIIALVYKGYKL
PLENGLKLKGSNDVKINDAPVMDGFKGTKEDDMIKALSEMTPEVRRYIAEVYAPSK
NKQRIELEFTTDLGVIGDITSLKMKYYPQMSQSLSRDSSGKLKTRGYIDLVS GAS
FIYRGNTSQSSSDKNVTKSSQEEQNAKEELQSVLNKINKQSSKN"
198761..200173
/gene="ftsA"
198761..200173
/gene="ftsA"
/note="SAV1185"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="BAB57347.1"
/db_xref="GI:14246955"
/translation="MEEHYVVSIDIGSSSVKTIIVGEKFNHGINVIGTGOTYTSIKNG
LIIDFOIARQAIDTIKASIASGVDIKEVFLKPLPIGITEVYDESNEIDFYETEING
SHIEKVLGRKERNDOVETEIVNFPRIIVDKNEVSDPKELIARHSLKVEAGVIAI
QKSLILNIMKVCBACGVDDLVDSYDAYNYGSIILTATEKELGACVIDIGEDVOTVAYE
RGELVNDASCEMAGROITDIAQGLANTSYETAEKVHQYGHAFYDSADQDITFEOV
DSDETVOYTKOLDSFTEARVEIIFEVFDVLQDLGLTKVNGGFIYTGGSANLLGVKE
LLSDMYSEKVR IHTPSQMGIRKPEFFSAISTISSIAFDELLOYVTINYHNDRETED
VIDDKDNESKLGGEFDWFKRKNKKDTHENEVESTDEEIQSEDNHQEHQNHQHVQ
DKDKDEESFKKLMLKSLF"
200206..201378
/gene="ftsZ"
200206..201378
/gene="ftsZ"
/note="SAV1186"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="BAB57348.1"
/db_xref="GI:14246956"
/translation="MLEFEQGFNHLATLKVIGVGGGGNNVNRMIDHGMNVRFIANI
TDQALNLSKAEKSIQIGEKLTGRLGAGANPEIGKAAEESREQIEDAIQAGDMVFYT
SGMGGTGTGAAPVAKIAKEMGALTGVGVVTRPFSFEGRKRQTQAAAGVEAMAAVDT

gene
CDS
201638..202429
/gene="SAV1187"
201638..202429
/gene="SAV1187"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57349.1"
/db_xref="GI:14246957"
/translation="MNDNFKKQPHHLIYEELLOQGITLIGITTRGDGLSDYFKPKNFMA
RYIDRPNYTOHOLAELAEIAPDKNMVPIQIOTHENKVCACITKDDTGTNIDTLDAL
FDGLADYDSNVLLTMCYACQVPEYFYSTKHHEIALAHAGMRCGTYTEIVKEVLKHVN
FDGLAHVIGPSTSSSYEINDDIKNKFEPLPIDSANXYETIETRGDRDRHIDLKKANAE
LNYGVKPENIYTAYATSEHLELFFSYRLEKGGTGRMLAFIGQQ"
202447..203121
/gene="SAV1188"
202447..203121
/gene="SAV1188"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57350.1"
/db_xref="GI:14246958"
/translation="MRVKDNLQOISTOINDKSEKNNESTKPNVTAVTKYVTIERAKEA
YEAIRHFGENRLEGFTQKKEALFSDAVIIHFISLQSRKKVDVINDVDYFHALDRLSL
AKEINKRAEHKICFLQVNVSGEASKHGIALDVDYDFIDDLKQTEIVGLMTMAPL
TDDEAYIRSLFKQLRLKKEETQRLNLEYAPCDELSGMSNDYLLIAVEEGATFVRIGTK
LVGEE"
203118..203681
/gene="SAV1189"
203118..203681
/gene="SAV1189"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57351.1"
/db_xref="GI:14246959"
/translation="MSHALKDLFSGFVIDDEEEVEVPDQKQOVNEAPAKEQSQOFT
KQNAIKSPQKSASRYTTTSEERNRMSNYSKNNSRNVTVMNNAITPNNASQESSKMCL
PEPRVSDTQDIADELKNRRATLVNQRIDKVSAKRIIDFLSGTVYVYAGGDIQIRVGTD
IFLCTPDNVEAGSITDHIENMEHSFD"
203693..203983
/gene="SAV1190"
203693..203983
/gene="SAV1190"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57352.1"
/db_xref="GI:14246960"
/translation="MDINVLATIPFKFILFVVEIYFGMIIFYFTSWPSTIRETKVGVF
LAKIYEFLOPFRKVIPIPIGIDISSIAAIIVLVLFQKGLIQFNWILLIQ"
204234..204872
/gene="SAV1191"
204234..204872
/gene="SAV1191"
/note="similar to cell-division protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57353.1"
/db_xref="GI:14246961"
/translation="MEVICGSYEDLVNSFYGGPNAERKRAIISPNYEPKESDFELTL
MEIYEPKFTVLKHQHILGLTMSLGIEREOVGDIIVNERQIOLVTRLSRELSFIMLELOR
IKGASVKLYTIPVTDMIQSNENKKNESATVSSLRDLVDVVKEMIRKSVTKIAKOLEKKR
VKYNHTIVDSADPQLOLANDLISIOGFGRAHITDLOGGKTKDKTHITTYRTLFK"
204896..205513
/gene="SAV1192"
204896..205513

/db_xref="SAV1192"
/note="similar to cell-divisio initiation protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57354.1"
/db_xref="GI:14246962"
/translation="MPFTPEIKNKESRVKNGLEPTEVANFLEQLSTEIERLKEDKK
QLSEKVEERTNIKSYQDVHVSQSDALIQAKAGEETKQAAEKQAEAIIAKAAQAQ
MVGDAVEKARLAFQTEDMKRQSKVFRSRMLVEAQLDLKKNEDWDYLLNYDLDAQ
VTLENLHLENDLKPDEVAANAQNASNPDDNQNSDSETTKK"
205734..208487
/gene="iles"
205734..208487
/gene="iles"
/note="SAV1193"
/codon_start=1
/transl_table=11
/product="ile-tRNA synthetase"
/protein_id="BAB57355.1"
/db_xref="GI:14246963"
/translation="MDYKETLLMPKTDTPMRGGILPNKEPQIQEKWDAEDQYHKALEKN
KGNETFLHDGPPYANGNLHMGHALNKLKDFIVRYKTMQGFYAPVYVPGWDTGLPTE
QALTKKGVDRKKSTAEREKCKEFALQLELOKOPRRLGVRCNDNDPYITLKPEVE
AAQIRIFGENADKGLIYKKKPYVSPSSSESLAEAEIEYHDKRSASIIYVAFDKDK
GVYDAAKFLIWTTPWTIPSNVAITVHPELKYQNVNGEKYIIABALSDAVAELD
WDAASKLEKEYTGKLEYVVAQHPFLDRESLVINGDHVTDTAGTCVHTAPGEGEDD
IYVHGQELPVIIPIDDKGVTEEGQFEGMFYDKAMKAVTDLITKGAALLKLDFTIH
SYGDWRTKPVIFRATPOWFASISKVRQDILDAIENTFKVNMKGTRIVNMVDRDGE
WVLSRQVNGVLPVFEYAENGLIMTKETVNHVADLFAEHGSNIWPEREAKDLLPEGF
THPGSRNGFTKETDMDVWFDGSGSHRGVETRPESLPADMYLEGSDDYRGWFNS
ITTSVATRGVSPYKFLDSHGFGVMDGSKMSLGNVIVPDQVVKQADIARLWSS
TDYLAURLSDEILKQTSQDVYRKIRNTLRFMLGNINDFNPTDSIPESELLEVDYRLL
NRLREFASTINNYENFDYNIYQEQVNFIVNELSNFYLDYGDIDILYIEORDSHRRS
MOTVQLVDMTKLAPILVHTAEVWSHTPHVKEESVHLADMPKVVEVDQALLDKW
RTFMNLRDQVNRALETARNEKVIGKSLEAKVTASNDKENASELTSFDALHOLFVVS
QVYVDKLDDQATAYEHGDIVIEHADGCKERCWCNYSDDLGAVDELTHLCPRCQQVVK
SLV"
208778..209575
/gene="SAV1194"
208778..209575
/gene="SAV1194"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57356.1"
/db_xref="GI:14246964"
/translation="MYHNSNANFVNGITLNVDRDKNELKPFYEDVLGLNIINETLSIQ
YEVQNNHVITVELONGREPLMSEAGLFHIAIKLQISDLANLLHLSEYDIPVNGG
IQPASLSLFEDEPGNGKFXFYVDKDEAQWTRQNDLVKIDIRPLNPLRVSHATKLWL
GIPDDAIIGALHIKTHLSEVKEYLDYFGLQESAYMDDYISIFLASNGYYQHLMNDW
VSATKRVENFDYGLAIVDFHPVETTHLNLQGPDGIVYRFNTHIEVED"
complement(209636..209848)
/gene="truncated-tnp"
complement(209636..209848)
/gene="truncated-tnp"
/note="SAV1195"
/codon_start=1
/transl_table=11
/product="truncated transposase"
/protein_id="BAB57357.1"
/db_xref="GI:14246965"
/translation="MLSFTRMSYRGINKVKRELGFVLMALNIRKIAAQRAVHYKIIHK
KADFYOIINRNOLFTLPKLNMSQPPS"
210399..210890
/gene="lsp"
210399..210890
/gene="lsp"
/note="SAV1196"
/codon_start=1
/transl_table=11
/product="lipoprotein signal peptidase"
/protein_id="BAB57358.1"

/db_xref="GI:14246966"
/translation="MHRKFIKFTSILIAVFFVFIQVQTKYIIATMTKIGDSFEVIPHF
INTSHRNNGAAGILSKKFFFTFIITIIILIAVYFFIKDAQNLNFMQVAISLIFAG
ALGNFIDRLITGEVVDFTDNIHFGYDFPIFNIAADSLTIGVIIIIALLKOTSNKKEK
EVK"
210890..211807
/gene="SAV1197"
210890..211807
/gene="SAV1197"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57359.1"
/db_xref="GI:14246967"
/translation="METYEFNITDKEQTGMRVDKLLPELNNDWSRNOIQDWIKAGLVV
ANDKVVKSNTKVLNDHIVVTEKEVVEADILPENLNDIYVEDDDVAVVYKPKGMVHV
PSPGHYNTLVNGLMYOIKNLGSIENGELRGIVHRIDMDTSELLMVAKNDIAHRLGLE
QLMDKSVRKYIALVHGNIPHDYGTIDAPIGRNKNDQSQMAVVDQDGEKAVTHFNVLH
FKDTYLVECEQLETGRTHQIRVHMKYIIGFPLVDPDPKYGPFKTLDIGGQALHAGLIGFEH
PVTGEYIERHAELPQDFEDLLDTIRKDA"
212207..212734
/gene="pyrB"
212207..212734
/gene="pyrB"
/note="SAV1198"
/codon_start=1
/transl_table=11
/product="pyrimidine operon repressor chainA"
/protein_id="BAB57360.1"
/db_xref="GI:14246968"
/translation="MSERIIMDDAAIQRTVTRIAHEILEYKNGTDNLILGIKTRGEY
LANRIQDKIHQEQRIPTGTIDITYFRDDIEHMSLTTKDAIDIDTIDTKVVIID
DVLYTGRTVRASLDAIILLNARPIKIGLAALVDRGHRELPIRADPVGKNIPTSKETVS
VYLEMDORNAVITK"
212952..214259
/gene="pyrP"
212952..214259
/gene="pyrP"
/note="SAV1199"
/codon_start=1
/transl_table=11
/product="uracil permease"
/protein_id="BAB57361.1"
/db_xref="GI:14246969"
/translation="MONDEFERTVKPVDVNEKPOPAQWAFLSLOHLFAMFGATVLV
PFLTGLPISAAALLASGIGTLLYILITKAQIPAYLGSSFAFTTPIITGLSTHSLGDMLV
ALPMSGMVYIIGILIKLSTANMLKLLPPVVGVPVIMVIGLSLAPTAVNMAYENPG
DMKGYNSFLIVAMITLLVTIVVQGFPGFLSLIPVLVGIIVGVYVAIFMGIVAFDAI
MSKARDIFPHIYLPFDKYDVPFSLGLVLMVPIVFTVSEHIGHWMLNKIVGRNFFE
KPLGDKSIIIGDGYSTMFASITIGGPPSTTYGENIGVLAITRIYSIYVIGGAATVILVA
FIKGTALISSIPTVMGVSYILLFGIIAASGLRMLVESKVDVFANNRLNVIASVILV
GIGNLVNKEIGINLQIEGMALAALSIIILNLILPKKKQNN"
214287..215168
/gene="pyrB"
214287..215168
/gene="pyrB"
/note="SAV1200"
/codon_start=1
/transl_table=11
/product="aspartate transcarbamoylase chain A"
/protein_id="BAB57362.1"
/db_xref="GI:14246970"
/translation="MNHLLSMHEHLSLTDIOIKLIKQASOFKSGERQLPNFEGKYVANLF
FENSTTKCSFEMAELKGLKTIISFSTSTSSVSKGESLYDTCCTLESIGDGLVIRHP
FNNYIEKLANINIPIANAGDGSQHPTQSLDLMFTIYEYEGYEGEGLNVLCGDIKNSR
VANSYTHSLKALGNWFNPNWADDSLEAPYNDIVDTETVDIVMLLRQHHERHGL
AETTFRAADDYHQKHGLNEVRYNKLQEHATVMPHPAPVNRGVETQSDLVEASKSRIFKO
MENGVYLRMAVIDELLK"
215186..216460
/gene="pyrC"
215186..216460
/gene="pyrC"
/note="SAV1201"

/codon_start=1
/transl_table=11
/product="dihydroorotase"
/protein_id="BAB57363.1"
/db_xref="GI:14246971"
/translation="MKLIKNGKVLQNGELQQADILIDGKVIKQIAPAEPSNGVDIID
AKGHVSGFVDVHVHLRPGGEYKETIETGKAAARGGFTVCMPNTRPVDSVEH
FEALQKLDDNAQVRVLPYASITTRQLGKELVDLPALVKEGAFATDDGVGQTASMM
YEGMTEAAKVKAIVAHCEDNSLIYGGAMHEGKRSELGIGIPINICISQIAROVLL
AEAAGCHVHVCHSVKESVRVIRDAKRAGIHVTAEVTPHHILLTDEDDIPGNNAIYKM
PPLRSTEDREALLEGGLDGTIDCIATDHAPHARDEKAOPMEKAPGICVGSETAFFELY
THFVANGDWTQQQLVDYLTIRKPETFNLEYGTLKENGYADLTIIDLSEQEIKGDEFL
SKADNTPFIVKGVGNPILTMVEGEVKFEGDK"
216462..217562
/gene="pyrAA"
216462..217562
/gene="pyrAA"
/note="SAV1202"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate synthase small chain"
/protein_id="BAB57364.1"
/db_xref="GI:14246972"
/translation="MQSKRYLVLEDGSFYEGYRGLSGDNLTVGEIVFNTAMTQYQTIS
DPSYTGQIITTYPLIGNYGINDDFESLVPTLNGIVVKEASAHPSNFRQOQTLHDVL
ELHQIPGTAGVDRSTRIRKIRQHKVAKAGTDRKEDIDQLVKHQQOVELPKNEVIVS
TKTPYSTGKDLVVVDFQGNVIRNLNVRGCNVTVVPTTAAEILMAPDGVML
SNGCPNVECAIPMIOGILGKIPFGICIGLHOLFALSOGASSPKMFKFGRGANHPV
KNLETKVDITSONHYAIDIDLSKSTDLEVTHLALNDGTVEGLKHKILPAFSVQYHP
EANPGPSNLTDFDFFVAMTNFKERKHINA"
217555..220728
/gene="pyrAB"
217555..220728
/gene="pyrAB"
/note="SAV1203"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate synthase large chain"
/protein_id="BAB57365.1"
/db_xref="GI:14246973"
/translation="MPKRNDIKTLIVIGSGPIIIQAAEFYAGTQACIALAKEGYRV
ILVNSPATIMDKIADKIETPLTHDFIAKIRKQEPDALLPLGGOTGLNMAIQ
HESGVLQNNVOLLGTSLTSIQCAEDREMFRLMNDLNVVPVSPESDIVNTVQAEKFE
QVGYPLIYRPAFTMGGTGGICCHNDEELHIVSNGLHYSPATQCLLERSIAQKFEIY
EVRKNDNAIVVCNMENIDPVGISIVVAPQSOTLSDEVYQMLRDVSLKVIKRALG
TEGNCVOLADPHSFDYIIIEVNPVRSRSLASASKATGYPIAKIAKIAVGLTIDEM
LNPITGTYAAFEPTLDVVISKIPREPDKPEKGBRELGTQMKATGEVMAIGRTYES
LLKARSLEYGVHHLGPNGESFDLDYIKERISHODDERLFIGEAIRRGITLLEEIH
MTQIDYFFLHKQNLIDIEHQLEHOGDLEYLYKADYGFSDKTLTAHRFNMTTEEYQ
LRMENDIKPVYKMDTCAAEFESSTPYGYETENESIVTDKEKILVLGSGPIRIG
GVEFYATVHAWAIQKAGYEAIIVNNNPETVSTDFSTSDKLYPEPLTEEDVMNINIL
EKPGVVQFGGQTAINLADKLAKHGKILGTSLNENARADEKFEALLRKINVPQ
QKGSATPEEALANAAEIGYPVVPVPSVVLGRAMEIVDNDKELNRYTQAVKASPEH
PVLVDYLTGKEIVDAICDGETVPIIGIMEHIERAGVHSGDSIAVYPPQTDELA
TLEDYIKLAKNLNLIGINIQFVIAHGVVYVLEVPNRSSSTVPFLSKTIDIPMAQL
MRAIIGETLDMGOYEQVQAEYGVKAPVPSFNKLNNDITLGPENKSTGEMAGKL
TTLKALEFGKITGSGVEVKDHLTVMTSDKDKKEWKLQAEKLVNGYKILATSGTAN
KLAEYDIPAEVVGKIGGENDLITRQNGDVQIVINTMTKGREVERDGFQIRRTTVENG
IPCLTSLDTANALTNVIESMTFMQM"
220838..221530
/gene="pyrF"
220838..221530
/gene="pyrF"
/note="SAV1204"
/codon_start=1
/transl_table=11
/product="orotidine-5-phosphate decarboxylase"
/protein_id="BAB57366.1"
/db_xref="GI:14246974"
/translation="MKDLPIIALDFESKEKYNQFLDLDSELSFVKVGMELFVQEGPOL
INEIKERGHDFLDKLHDIPTNYGKAMEGLAKNLVDLVNHAAGVYKMMSEAIKGLR
KNHQTKIIVTQLTSTTEDMLRHEQNIOTISIEEAVLNAYKLANAAGLDGVVCSPLS
RMLTEKLGTSFLKVTPGIRPKGASQDQHRITTPPEARQLGSGTHIVVGRPIQTSDNPV

ESYHKIKESWLV"
221530..222141
/gene="pyrE"
221530..222141
/gene="pyrE"
/note="SAV1205"
/codon_start=1
/transl_table=11
/product="orotate phosphoribosyltransferase"
/protein_id="BAB57367.1"
/db_xref="GI:14246975"
/translation="WAKEIAKSLLDIEAVTLSPNDLYTWSSGKSPDIYCDNRVTGLGP
LYVRGIRGLINLRIKEHFEVEVISTGATAGIPHAAFTAEKLGKLPNNYVRSSNKSHGK
QNOIEGAKSEGGKVVVIEDLISTGSSSVTAVEALKQAQAEVLGVVAIFTYGLKKADDT
FSNIOLPFVTLSDYNELIEVAENEGKISSIEDIQTILVEWRDNLA"
222171..222383
/gene="SAV1206"
222171..222383
/gene="SAV1206"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57368.1"
/db_xref="GI:14246976"
/translation="MNDKTSNDLYGKIKHCNEFINHSNDSNLSSSHVDVDESSTKQKHI
KNKTIIDHNDLDFKHVKDILRRQGQI"
222820..223221
/gene="SAV1207"
222820..223221
/gene="SAV1207"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57369.1"
/db_xref="GI:14246977"
/translation="MDIPKITTFMFNKAERAVKLYTSLFDESELIITMAKYGENGPG
DRGTQVLSITLNGQVFWAIDANSGETELPMNPAISLFTVTKDTIEMERLFLNGLKDEGA
ILMPKTNMPPYREFAWQDKFGVSQALPE"
complement(223484..225181)
/gene="SAV1208"
complement(223484..225181)
/gene="SAV1208"
/note="similar to fibrinogen binding protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57370.1"
/db_xref="GI:14246978"
/translation="MAYDGLFTKKMVESLOFLTTRGVHKKINQPDNDTILMVYRQNRQN
HOLLISIHFNFSRLQTLTKYDNPFPNPFARVFRKHLKEGGIESIKQIGNDRRIEID
IKSDEIGDTIVRTVLTLEIMGKHSNLI.LVDENRKIIIEGFKHLPNTNHYTVRPGFNY
EAPPTQKINPYDITCAEVLKYIDFNAGNIKAKOLLNOFEFSPLITINEVSRQPMFS
STLPEAFDEVMAETKLPPPTPIEKNHETQCKEYFIKLNQFNDPTVYDSLNDLDRF
YDARGERKVRANDLRFVQOQLHKYQNKLAKEIEYEQESKNKDETEQLYGLELITAN
LYRIKQGEFTALNYTNEEVYIPLNPTKPSANAQYIKQYKNRMTKRELEQHOIQ
LTKDIDYFSTIEQQLHHISVDHIDIEIRDELAEQGFMKQRKQTKKKKQIOLQHYVS
TGDDDIYFGKNNKQNDYLTNRKAKKTHWLTHTKDIIPGSHWTFNDAPSOTTIKEAAML
AGYFSKAGNSGOIPVDYTLTKNVHPKSGAKPGFVTDNQKTLVATPDVELIQMKQL"
225456..226079
/gene="gmk"
225456..226079
/gene="gmk"
/note="SAV1209"
/codon_start=1
/transl_table=11
/product="guanylate kinase homolog"
/protein_id="BAB57371.1"
/db_xref="GI:14246979"
/translation="MDNEKGLLIVLSPGSGVGKTVRRKIFEDPDSYKYSMTTPO
MREGVDGVDFYFKTFRDAFEALIKDDQFTYEAAYVGNVYGTVPVYVKDFTMDSGHDVFL
ETEVEGAKVRRKFPFDALFIFLAPSLDLHLRELVRGFTESNEKIQSRINERARKEVEM
MNLVDYVVVNDVEVLEAKNRKIQICIVEAHLKREVEREAKYRKMILEAKK"
226079..226297
gene

CDS

/gene="SAV1210"
226079. .226297
/gene="SAV1210"
/codon_start=1
/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57372.1"
/db_xref="GI:14246980"
/translation="MLNPPLNQLTSQISKYLIATTAARAREIDEOPETELLSEYHS
FKPGRALAEIADGKIRPVISSDIYGR"

gene

226513. .227712

/gene="SAV1211"

226513. .227712

/gene="SAV1211"

/codon_start=1

/transl_table=11

/product="pantothenate metabolism flavoprotein homolog"

/protein_id="BAB57373.1"

/db_xref="GI:14246981"

/translation="MKKILLAVTGGTAAYKAIDLTSKLTOSGYEVRVMTNHAQKFTV
PLAQAISNAVYTDPIENPSEIQLHAGWDADIIVAPATANTIAKLSVGIADDL
VTSLIATEPKFTAFAMNVHMYENKRTOQNIILKEDGYHFIPEGSGFLACGYVARG
RMEEPQIVSVIDAHFQNSNRLANSFQDRALVATGPTIEVDPPVRFVNRSSGRMG
YATAEALNRGAIVTLVAGPTITLEDPKDIEIVHQSAEEMFEQVTSRFDEQDIIWRAA
AVSDYTPVDVLEHKMKDGLSVSFKRTDILKYLGEHKTSOYLLIGFAAETEDIENY
AQOKLRKNADVIIISNNGVDMSIGSFSSDDNLTMHIFKNNEKVNKKKKVYVLAQIIL
ELETRQ"

gene

227712. .230120

/gene="pria"

227712. .230120

/gene="pria"

/note="SAV1212"

/codon_start=1

/transl_table=11

/product="primosomal protein"

/protein_id="BAB57374.1"

/db_xref="GI:14246982"

/translation="MIAKVIDVASKSDYKEDYIIPQELSEIVQGRVIVPFGPRT
IQGVMEVTAEPDAQLDVKLKKLIEVKDIOBELTSELIALSEWMSGTVIRKISMLE
VMLPSAIKAKAFKMKODIELPSALLQKDFKHGYIKQYDKAKNDIOLMLKKLDD
IIVEKTLITONITTKRAVRVIEGYHDPDEVLAKLEKVIKQYDLYALSEEQHKTIFL
TIDEDGFSKSLDGLIKGYVEKYDAVVERPDKDFVEQESKQOLTEDQYKAYEAI
KAKIVSQEQTFLHGVGTGKTEVYLOTEIDVLSQKQAMMLVPRIALTPOWLREK
RRGDVAVLHSLGNGERYDEWOKIRDDRARVSGARSVEAFPKNLGLIIIDEHE
SNYQEDYPRYHAREIAQRSEYHHCPIVLGSATPCLESYARAEKDVYHLLSLPVRN
QOALPEIDIVNREELSEGNSMFSKDLREAQOLDRQEVVLFUNRGYASFMLCR
DCGYPCPCNDISLHYHTKTLDDLKCHYCYQETPPNQCPCSEHIRQVGTQKVE
ELIQOEFEDARIIMDVDTTSKGAHEKLTIFEKNGDILLGTQMIAGLDYPNITL
VGLNADTMLNPDRASERTYQLTQVAGRAGREKAGOVIIOTNPDHYSLTDVOK
NDYLFYROEMEVRLGKYPYVYLINFETISHKEMKVMQEAQSHVHKILLQHLTEKAL
VLGSPSAAALARINNEFRFQILLYKSEPGLLQAIQIQLDDYYHEKFIKEKALKLIDIDP
QMMW"

gene

230626. .231579

/gene="SAV1213"

230626. .231579

/gene="SAV1213"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57375.1"

/db_xref="GI:14246983"

/translation="MKTLGCLLLIMLVVAGCSFGGNHLSKSKSEESKQETVKES
EEKDPDLKYEIEKKMKGIKAPSLDKLDPLMTEKSTNSKGLQGWKDYKELMGKV
ELADYRFTKDSGSSIKQVDAPFKGKGLKRVKVIETHDDVKQVDYVDPDGKKIGNS
NTPVYELMTYIKQGLVYASVEGVSIVYHKDDAIKYDDYDKLKLKLSQLT/KLDHPK
VPYVAQIKFQCVLTSVFMTHGSKDTKDEVLPALAYFTFSPKNYEDKSNPDPRVLN
LVHMDFLNASSDFGNAHFVLSKYIKVESNYETASDLSK"

gene

complement(231742. .231945)

/gene="SAV1214"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57380.1"

/db_xref="GI:14246988"

/translation="MITAEKKKKKFLPNFQKSIYSLRDEMOMNLVEGQOQKFRAK
QIFELVYLRHDSIDMTNLSKDLQKDNFTVTTLTVVKQESKDGITKFLFELQD
GYTITVLMRHDYNSVCVTVQVGRIGCTFCASTLGLKRLNLEAGEIVSVLTQOKA
LDATERVSOIVIMGIGEPFENYDEMDFLRVNDNSNLNIGARHTVSTSGIIPRIY
DPADEDIOINFAYSLHAKKDEVRSLRNLNRYNAYNVEKLEIAQYQOEKNRRVFEYG
LFGGVNDQLEHARELAHLKGLGNLVNLPVNVHVPERNYVTKANDIFKPEKELKRLG
INATIRRSQSDIDAAACQOLRAKERQVETR"

gene

236068. .236811

/gene="SAV1219"

236068. .236811

/gene="SAV1219"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57376.1"

/db_xref="GI:14246984"

/translation="MQVKNKVIYILLALFLSGFIHKFYAGKNMOGILHLIFCWTGIPH
ILAIISAVIVFKPADEQGNVTL"

gene

232244. .232732

/gene="SAV1215"

232244. .232732

/note="SAV1215"

/note="similar to polypeptide deformylase"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57377.1"

/db_xref="GI:14246985"

/translation="MAIKKLVPASHPILTKKAQAVKTFDDSKRLQLQDLEDTMYAQEA
AGLCAPQINOSLOVAIIDMEMEGLQLVNPKISQSNETITDLGSGTLPDVGVEVIR
SKMIVESYDVNGKNKVELTAHEDVARMILHIIDQMGPPTERRADRLITDKEVAYFI
ND"

gene

232725. .233660

/gene="SAV1216"

232725. .233660

/gene="SAV1216"

/codon_start=1

/transl_table=11

/product="methionyl-tRNA formyltransferase"

/protein_id="BAB57378.1"

/db_xref="GI:14246986"

/translation="MTKIIIFMGTPDFTVTVLEMLIAEHVDVIAVVTOPDRPVGRKRVMT
PPVKKVAMKYDLPVYQPEKLSGSEELQLQDLVDLIVTAAGQLLPESLLALPRIG
AINVHASLIPKRYGGAPIHOAIIDGQETGITTMYVVKLIDAGNIISQQAIKTEENDR
VGTMDKLSVLGADLLKETLPSIIEGTNESVPQDDTQATFASNIRREDERISWNKPCR
QVFNGIRGLSPWPVAYTTMDDTNLKIYDAELVETKNINEPTTIETTKKAIIVATNDN
EAVAIKDMOLAGKRLAANYLSGAQNTLVGKKLI"

gene

233579. .234964

/gene="SAV1217"

233579. .234964

/note="SAV1217"

/note="similar to RNA-binding Sun protein"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57379.1"

/db_xref="GI:14246987"

/translation="MSWERENVSQCLFKWCAKHTSREETYMIENVRSLAFDTIQDILN
EGAYSNLRINEVLSNELNAMDKALFETIYGVTKRYTLDLYLKPEVTKIKAWVRQ
LWMSIYOYVLDKVPNHAIINEAVEIAKERGGYHNGVNVNGILRTMPBSDLDFENEI
ADPKPMALEYSNPKWIIDHWATHYGLEETETILQSLETTSTTVRANLTRASLDDLI
EKLQDEGYDVERDHDLPYCLHIGGQPIIHSRPFKDFGVSIOQKSSFMFAHMMVNRHD
HVLDCASPGGKACHIAEVLMPGQVDASDIHDHKIDLINFNKIKRLNLNIKAFQHDA
TKPYDKTYDKILVDAPCSGLGVNRHKPEIKYQSKOHIISLVELQLEILENKNYKI
GGEIYCTCTIEQLENNVIYTLKNNKNFEPEFQHPITGELVKTLQIMPQDFNSDG
FFITKIKRKDN"

gene

234967. .236061

/gene="SAV1218"

234967. .236061

/gene="SAV1218"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57380.1"

/db_xref="GI:14246988"

/translation="MITAEKKKKKFLPNFQKSIYSLRDEMOMNLVEGQOQKFRAK
QIFELVYLRHDSIDMTNLSKDLQKDNFTVTTLTVVKQESKDGITKFLFELQD
GYTITVLMRHDYNSVCVTVQVGRIGCTFCASTLGLKRLNLEAGEIVSVLTQOKA
LDATERVSOIVIMGIGEPFENYDEMDFLRVNDNSNLNIGARHTVSTSGIIPRIY
DPADEDIOINFAYSLHAKKDEVRSLRNLNRYNAYNVEKLEIAQYQOEKNRRVFEYG
LFGGVNDQLEHARELAHLKGLGNLVNLPVNVHVPERNYVTKANDIFKPEKELKRLG
INATIRRSQSDIDAAACQOLRAKERQVETR"

gene
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57381.1"
/db_xref="GI:14246989"
/translation="MLEAQFTDTGHRDKNEDAGFIYVQNTQOOLLVLCDMGGHKA
GVAKFVDELKSRFEAEHLIEEHAENLRNKNIDINFOLYHYAENAEYKGMGTT
CVCAIVFESVIVANYGDSRAYVINSRQIEITSOHSFVNLVLTGOITPEEAFHPQ
NLITIKVMGTDKRVSPDLFIKRLNFYDYLNLSDGLTVYKDNKIEKRLLVKEGTIEDH
GQILMQLADNHSKONVTEILAAIEGDKV"
236808..238802
/gene="SAV1220"
CDS
236808..238802
/gene="SAV1220"
/codon_start=1
/transl_table=11
/product="protein kinase"
/protein_id="BAB57382.1"
/db_xref="GI:14246990"
/translation="MIGKIINERYKIVDKLGGGMSTVYLAEDTILNIKVAIKAFIP
PREKETLKREREVHNSQLSHONIVSMIDVDEEDCYLVNXYEGPTLSEYIESH
GPLSDVTAINFNQLLDGIKHAHDMRIVHRDIKPNILIDSNTLKIPEFGIALSE
TSLTQTNHVLGTVQYFSPQAKGEATDECTDIYSIGVILEMLVGEPPNGETAISIA
IKHIQDSPNVTTDVKKDIPQSLSNVILRATERKDRANRYKTIQEMKDDLSVLHENRA
NEDVLEDMKMTIAPVLPKKEDIAKHISEKSNQKRETTQVPIVNGPAHHQOQKPEG
TVYEPKPKKSTRKIVLLSLIFSLMIALVSVAMAFNGKYEETPDVIGKSVKMPNV
IFNKNMLGKISRSYDKYPENELIKTTPNTGERVGDSDVNVISKGPVKMPNV
IGLPEEALQKLKSLGDKDVTIEKYVNOAPKGYTANQSVTANTEIAIHDSNIKLYES
LGIKQVYEDFEHFKSKAKKALEKGFVESKEEYSDDIDEGDVISQSPKSGVDEG
STISFVSGKSDSDSVKTTTESVDVPTYGKNDKSQKVYIKDKNDGSGTERKGSFD
ITSQRIDIPRIERIEKGTASYIVKVDGKTVAEKEYSYDDI"
239030..239905
/gene="SAV1221"
CDS
239030..239905
/gene="SAV1221"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57383.1"
/db_xref="GI:14246991"
/translation="MKTGRIVKISGVYQVDVNGERFNTKPRGLFRKKKFSFVWGDI
EFEVQNINGYIHQFERKNELKRPVSNIDTLVIVMSAVEPNTSQLDRDLVIAHS
YOLNARVLTKKDKTPIEKQFENELLKIYENIGYETEFIGNDDRRKIVEAWPAGLI
VLSQSGVGKSTFLNHYRPLENLETNISKSLNRGKHTTRHVELFERQNGYIADTPGF
SALDPDHIDKIDKDYFLENLRYGETCKFRNCNHIKEPNCNVKHQLEIGNIAQFVYD
YLQLENEISNRKVY"
239906..240550
/gene="cfxE"
CDS
239906..240550
/gene="cfxE"
/note="SAV1222"
/codon_start=1
/transl_table=11
/product="ribulose-5-phosphate 3-epimerase homolog"
/protein_id="BAB57384.1"
/db_xref="GI:14246992"
/translation="MTKLYPSLLSVDFLDLOHELKLEAGVDGVHFDVMDGQVPMI
SIGLPILDVAVRGTTLPIDIVHLMENPEKYIASFHEGADMISIHVESTPHIHALQIM
IKHLDKAGVINPGTPIQIEPILDIVDVLVMTVNPFGVGSGQSIDQCVEKIAGLNA
IKMERQLNFDIEVDGSGVNTDTAKVCVENGATMLVTGSFFFKEDYKKVTOALKG"
240557..241198
/gene="SAV1223"
CDS
240557..241198
/gene="SAV1223"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57385.1"
/db_xref="GI:14246993"
/translation="WHINLLCSNRHLPDQIMAKSNEGKGGVDVKGALILLKHLQIIPFF
SVGDEDSVKEERQLLEQOLKPVQAEKADTDLALADKAVALGSDSITIIYGATGGR
LDHFFCAIQLLKRAYIKHVDHIEVDQNKTELPLPKGQHTVEKDKSYPIISFIPMTD
DVELSLAGPKYNLARQMLNIGSTLTISNEIESLQAKVTVHDDLILQIRSTDLEN"

gene
CDS
complement(241579..241767)
/gene="rpmB"
complement(241579..241767)
/gene="rpmB"
/note="SAV1224"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L28"
/protein_id="BAB57386.1"
/db_xref="GI:14246994"
/translation="MGKQCFTVTRKASTGNRRSHALNSTKRRWNAMLQKVRILVDGKP
KKVWSARALKSGKVTNV"
242210..242584
/gene="SAV1225"
242210..242584
/gene="SAV1225"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57387.1"
/db_xref="GI:14246995"
/translation="WTEISNDYGKIDISNEVIASVVGKAVECYGIVGMASRQQVRD
GIAELIGHENYAKGIKVTENNGVVDIDMTIIVSYGKISEVANNVOSTVYKLTLEKSLN
VSVNSINIYVQGVVRNNTGKA"
242599..244245
/gene="SAV1226"
242599..244245
/gene="SAV1226"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57388.1"
/db_xref="GI:14246996"
/translation="MISKINGKLFADMIITQAGNLSNADLVDSLNVYVPPDDGTGNT
MNLTSGREEVNNLSKNIGELGTFSGLLMGARGNSGVILSQLRGFCRNIESES
EINSKLLAESFQAGVETAYKAVMKPVGEGTILTVAKDAQAATKANNTEDCITELMEYI
IVKANESLENTNLLAVLKEVGVDSGGKGLLCVYEGFLKALKGERKEAKVAKIKDE
FVDEHDHPHGVNTEDIYGYCTEMVRFGKKNKKADEFQDMQSOFGDSLLVINDE
EIVKVHVHTPEKGVNFYNGQXGELLKLVANNRQHRREVIRKQHTAKPKWETVTA
IITISMGEGISELFPKSMGATHIISGGQTNPSTEDIVKVEQSKCKRAILLPNNNKIL
MASEQAASIDAEAVVPTKSIQGISALFQYDQVDATLEENKAQMSADSVNNVKSGLT
YAVRQTKIDGVEITKIDAFMGLIEDKIVSQSDQLTFTVTELLNEMLAEDSEILTVITIG
DRAQVATDNMINWIEEQYDPDVEVEHGGQPIYQYFFSVE"
244435..246495
/gene="recG"
244435..246495
/gene="recG"
/note="SAV1227"
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA helicase"
/protein_id="BAB57389.1"
/db_xref="GI:14246997"
/translation="MAKVNLIESPYSLQLKGIGPKKIEVLOQLNIHTVVEDLVLYLPT
QYLVKKKIELNQTITVKGKNRVKQIEITGNRVFNSQGTQTOENADVOLEPVRIKEG
IKOKIORDIQRALANDVTIHEMLTDELREKYLETDLFTLTHHPKSEKEDILLARPT
YAFTELTPELRMQMLNRLKESDDEAIEDYDLQVKSFIDRLPTELTAQKSSVNEI
FRLDKAPIHMRLLQDQSGSKTVAACICMYALKTAGYQSALMVPELTILAEQHAESLM
ALFQDSMNVALLTGSYKGRKRIILLEQNGTIDCLIGTHALIODDVIIPHNYGLVITD
EQHREGVNORLLREKGMATNVLTATPITPRTLAISVFGEMDVSSIKOLPKRKPPII
TTWAKHEQYDKVLMQNTSELKKGROYVICPILESSEHLEDVQNVVALYESLQQYVYG
SRVGLLHGKLSADEKDEVNQKFSNEIDVLVSTTVVEGVGVNVPNTFMNIYADPRFGL
STLHQLRGRVRSQQQSCYCVLIASPKETIERTIMTQTTDGLSELDLEMRGPGD
FEGVKQSGLPDFLVANLVEDYRMLEVARDEAAELIQSGVFFFTYQHLRHFVEENLLH
RSFD"
246713..247270
/gene="SAV1228"
246713..247270
/gene="SAV1228"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

```
/protein_id="BAB57390.1"
/db_xref="GI:14246998"
/translation="MKLKKRRRAIRQQIDSNFFITDHLSDLFQVSIQIRLDRTY
LNIPELFKRKLVAEKNDYQISSIEEQEFGDLQVNPVNVKAQSIDLITSDSVPHKTG
IARGHVLQANSLCALIKOPTVLTTHESSTQTEKVKLNDTVRAEARVNVQTAHYV
VEKSYVKHTLVFKGNKFMFYDRG"
247275..248261
/gene="plsX"
CDS
247275..248261
/gene="plsX"
/translation="SAV1229"
/codon_start=1
/transl_table=11
/product="fatty acid/phospholipid synthesis protein"
/protein_id="BAB57391.1"
/db_xref="GI:14246999"
/translation="MVKLAIDMMGGDNAPDIVLEAVOKAVEDFNKLEIMLFGDEKKVN
LNHERIEFRHCSEKIEDEPVRRAIKRKDDSMVMKAEAVKSGEADGCVSAGNTGALM
SVGLFIVRTIKGVARPALVYTLPTIDGKGFVLDVGNADAKPEHLLQYLAQGDYIYV
KIRGINPKISLNLNIGTEPAKGNLSYKSPFELLNQDHSLNFNVTAKTLMDGDTDV
VTGTYGNMVLKNGLEGTAKSIGKMLKDTIMSSSTKNKLAGAILKKDLAEFAKKMDYSEY
GGSVLGLEGTVVYKAHGSSNAKAFYSAIROAKIAGEQNIQVTMKETVGSNE"
248254..249180
/gene="fabD"
CDS
248254..249180
/gene="fabD"
/translation="SAV1230"
/codon_start=1
/transl_table=11
/product="malonyl CoA-acyl carrier protein transacylase"
/protein_id="BAB57392.1"
/db_xref="GI:14247000"
/translation="MSKTAIIPGQAKVQMAQDLFNNNDQATEILTSAAKTLDFDI
LETMTFDESGKGETENTQPALTHSSALLAALKNLPDFTMGHSGEYSSVLAADV
SFEDAVIVKRGQMAQFPPTGQSMVAVLGLDFDKVDEICKLSLSDDKIIEAPAN
CFQIYVSGKHALIDELVEKGSGLGAKRVMLAVSGPFHSLMKVTEEDFSSVINQFE
WRDAKPPVQVNYNAOGETDKEVTKSNMKQLYSPVQFINSTEWLIDQGVDFHFEIGFG
KVLGSLIKINRDVKLTSTQITLEDYKGNEND"
249167..249907
/gene="fabG"
CDS
249167..249907
/gene="fabG"
/translation="SAV1231"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-(acyl-carrier protein) reductase"
/protein_id="BAB57393.1"
/db_xref="GI:14247001"
/translation="MKMTKSALVTGASRGIRSGIALQLAEEGVNVAVNYAGSKEAEA
VVEETKAKGVDSFAIQANVADADEVKAMIKEVVSQFGLDVLVNNAGITFDNLLRMK
EOEMDDVIDTNLKGVCNCKQATPQMLRORSAGIINLSSVGAAGNPGQANIVATKAG
VIGLTKSARELASRGITVNAVAPGFIVSDMTDALSDDELKEQMLTQIPLARFGQDTDI
ANTAFPLASDRKAYITGQTIHVNGGMY"
250342..250575
/gene="hmrB"
CDS
250342..250575
/gene="hmrB"
/translation="SAV1232"
/codon_start=1
/transl_table=11
/product="HmrB protein"
/protein_id="BAB57394.1"
/db_xref="GI:14247002"
/translation="WENEDFKVDIIVDRIGLVADKVTEDASPKDDLGDADSLDIAELVM
ELEDEFGEIPEDEAEKINTVGDVAKFINSLEK"
250691..251422
/gene="rnc"
CDS
250691..251422
/gene="rnc"
/translation="SAV1233"
/codon_start=1
/transl_table=11
/product="RNase III"
```

```
/protein_id="BAB57395.1"
/db_xref="GI:14247003"
/translation="MSKOKKEIIVNRFRKRFDTKMTGLGTYQNTIDLYQQAESHSSFI
NDNNRLDHNRELEFEGDVALETVLSVLEDKHPNLPEGNLTNRATIVCEPSLVIF
AKTGLEMLLVGGEETGRTSPSLSDFAEFAGIYLDQGLDIIYVVKFAEKVIFP
HVEQNLGGVDFDKTFQOEYHQOQNGDVTYNLTKEGPAHRLFTSEVILQGEAIAE
GKRTKKESEQRAAKSAYKOLKQIK"
251569..255135
/gene="smc"
CDS
251569..255135
/gene="smc"
/translation="SAV1234"
/codon_start=1
/transl_table=11
/product="chromosome segregation SMC protein"
/protein_id="BAB57396.1"
/db_xref="GI:14247004"
/translation="MYYLKSIDAIFKFSFADQTNVQFDKGVTAIVGPNSGSKSNITDA
IKWVLGQSASLRGSKMEDIIIFSGAHRKAQNTAEVQLRLDNHSSKILSDVENEIVIT
RRLYRSGSEYYINDDRARKLDIADFLDSGIGKAEYSIIISQGRVDELTNAKPIDRRQ
IIIESAGVLKYKKRAESLNKLDQEDNLTRVEDILYDLEGRVPLKPEAAIAKEYKT
LSHOMKHSDIVTVHDIDQYTNDRQLDQRLNDLQGOQANKAEADQRLSQOIQQYKKG
RHOLDNVESLNTQLVKATEAFKYTGQLVLEERKNQSETNARYEBEENLMELLE
NISNEISEAQDTYKSLASKOKELNAVIRELEEQLYVDEAHDEKLEETKNYYTLMSE
QSDVNDNDRIFLKHTIENEAKSKRLSLRLEVFEQLKDIOQIKTKKYEYTFPENG
AVDEIKNIKEDLTDTKKQANEYEKLYQARYTERKMKTRIDSLATQEEYTYFPNGV
KHILKAKNKELGIHGAVAETIDVPSKLTQAIETALGASLQHVIVDSEKDGQALQFL
KERNLGRATFLPLNVIQSRVVATDIKSIAKEANGFISIASEAVKVAPEYONILGNL
NTIIVDHLKHANELARAIFYRTRIVTILEGDIVNPGSGWTGGARKSKSILSOKDELTT
MRQLEDYLRQTESFEQQFKELIKSDLSLEYFEKSKHNHTLKEQVHHFEMELDRLT
TQETQKNDHEEFEEKNDGYTSDKSRQTLSEKHTLESIKASLRLEDETERYTKLS
KEGESYTKQOTLHOKQSDLAVVKERIKTQOQTIDRLNQSQOTKHQKDVKEKIAF
FNSDEVMEGAQONIKQINGQOFTRRLSDELQKQRELNEQIDAOEATLQVCH
ODILATENYODIKAEQKIDVLIIHAIHLDNDEYQLTVERAKSEYSDSESIDALRK
VKMKSEIDALGPNVNAIQEPFELNERYTFLEQRTDLKAKETLSQIISEMDOEVTI
ERKEFTHALQGHFTAVFKQFGGDRLEQLTETADYLTAGIDIVVQPPGKKLOHLSLL
SGGERALTAIALLFAILKVRSAFPVILDEVEAALDEANVIRYAKYLNLSDETQFIVI
THRGTMFEAFDRLYGVTMQESGVTKLVSVALNTIDDYVLEEQ"
255135..256385
/gene="SAV1235"
CDS
255135..256385
/gene="SAV1235"
/codon_start=1
/transl_table=11
/product="signal recognition particle"
/protein_id="BAB57397.1"
/db_xref="GI:14247005"
/translation="MSFFKRLKDKFATKNENEVKSIREEQGDQKLEDTHSEGSTQDV
NDLAEANVKKPRKLESEADFDDGLLISIEDFEEIAQKMGAKFKAGLEKSRQNFQ
LNNLIARYRKVDDEFFEALEEMLITADGVNTVMTLTTEELRMEAQRRNIQDTELRV
IVEKIYIEIHOEDDSEAMNIEDRLNVLIMVGVNGVGTITTKGLAYRKMEGKKVM
LAAGDTFRAGAIQDLKRWGVRGVDVTSQSGDPAVMYDAINAAKNKGVDILICDT
AGRLQNTNLMOLEKVRVINRAVPDAPEALLCDATTCGNALSOARNFKEVTWNTY
GIVLTKLDCGTAKGGIVLAIRNELHIPVKYVGLGEQLDDLPFPNPESTVYGLFADMIEQ
NBEITTVENDQIVIEEKDDNHGAK"
256372..256704
/gene="SAV1236"
CDS
256372..256704
/gene="SAV1236"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57398.1"
/db_xref="GI:14247006"
/translation="MGQNDLVKTLRMNLYLDFYQSLLTNKQRYLLEFYLEDYLSLEI
AOTFNVSQAVYDNIIRTCGLVDYEDYEKKLELYQKFPQRREIYDEMQRHLSNPEQIORY
IQOLEDE"
256730..258097
/gene="ffh"
CDS
256730..258097
/gene="ffh"
/translation="SAV1237"
/codon_start=1
```


AIPNFGONNDEBEPPTTEIKTKRSEIKRQLEGGKLEKEKVRKIVKQDPGALGMLGTN
QNOQMOKMQLMPKKVREVAVETARKILADSYADELIDQESANQEALEAFQMG
IFIDEIDKVTNNHNSQDVSROGQVQDILPILGSGVIQTKYGVNTEHMLFIFAGAF
HVKSPSLIPELOGFPIRVELDSLSDVEFVRILTEPKLSLKOYEAALLQTEEVTVNF
TDEAITRLAETAYQVNDTNDIGARRLHTILEKMLEDSLSEAPSMNPAVVDITPQYVD
DKLSITNKDLSAFIL"
278505. .279278
/gene="cody"
278505. .279278
/gene="cody"
/note="SAV1255"
/codon_start=1
/transl_table=11
/product="transcription pleiotropic repressor"
/protein_id="BAB57417.1"
/db_xref="GI:14247025"
/translacion="MSLSKTRLENTLLOKHGIADVDFKQVADPOTISSVTVTNFIVSR
RGKILGSSNELLSQRIOMLERHIPSSEYTERLMEVKQTESNIDNVLTVFPPEPN
RELFIDSRTTIFPILGGGERLGLVLGRVHDDFNENDLVIGEYAAVIGMEILREKHS
EVEKEARDKAATWAINLSYSEKEATEHIFEELGGTEGLLIASKVADVADRVGITRSVIV
NALRKLESAGVIESRSLGMRGTFIKVKKEKFLDELEKSK"
279625. .280392
/gene="rps8"
279625. .280392
/gene="rps8"
/note="SAV1256"
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S2"
/protein_id="BAB57418.1"
/db_xref="GI:14247026"
/translacion="MAVISMKQLLEAGVHFCHOTRRNPKMKVYIFTERNGIYIIDIQ
KTVKKVDEAVNFKQVEDGGQVLFVGTKKQAOESVKSSEARAGQFYINORWGLGLIT
NYTKTIRIKRIRISEIERKEWEDGLFVLPKREVVVELKKEYDLRIKFLGSIIDKMSMPQA
LFVVDPKENATAEARKLNIPIVIGIYDNCDDPEIDYVIPANDDAIRAVKLLITAKMA
DAILEGQGGVSNNEVAAEQNIDLDEKEKSETEATEE"
280574. .281455
/gene="SAV1257"
280574. .281455
/gene="SAV1257"
/note="EF-TS"
/codon_start=1
/transl_table=11
/product="elongation factor Ts"
/protein_id="BAB57419.1"
/db_xref="GI:14247027"
/translacion="MATISAKLVKELRKTGACGMDCCKKALTETDGDIDKAIDYLRK
GIAKAAKADRIAAEGLVHVHETKGNDAIVIEINSETDFVARNEGFQELVKEIANQVLD
TKAETVALMETTLTPNGKSDVERIKEAISTIGEKLSVRRFAIRTKTDNDADFAYLHMG
GRIGLVTVEGSTDEAARDVAMHIAAINPKYVSEQVSEEEINHEREVLKQOALNEG
KPNIVKVMYEGRLRYLQETCAVDQDFVKNPDTVVEAFLTKGCKLVDFVRYEVGEG
MEKREENFADEVKGOMK"
281592. .282314
/gene="smbA"
281592. .282314
/gene="smbA"
/note="SAV1258"
/codon_start=1
/transl_table=11
/product="uridylylate kinase"
/protein_id="BAB57420.1"
/db_xref="GI:14247028"
/translacion="MAQISIKYKRVVLKLSGEALAGEGFGINPVIIKSAEQVAEAVAK
MDCETIAVIVGGNIWRGKTGSLDGMGRGTADYMGMLATVMNALODLSLEQLDCTRV
LTSTEMQVAEPYIRRRAIRHLEKKRVYIEPAAGIGNYFESTDITTAALRAAEVADVIL
MGRNNVGVYSADPKVKNDAVKYIEHLTHIQMLEGVLQYVMDSTASSFCMNNPIUTVFS
IMEGNTIKRAVMGEKIGTLITK"
282333. .282887
/gene="frr"
282333. .282887
/gene="frr"
/note="SAV1259"
/codon_start=1

/transl_table=11
/product="ribosome recycling factor"
/protein_id="BAB57421.1"
/db_xref="GI:14247029"
/translacion="MSDIIINETKSMOKSITESLRELANISAGRANSNLLNGVTVDVY
CAPTPVQOLASINPEARLLVISPYKTSVADIKAIIAANLGNVPPTSDEGVIRIAPV
ALTERRERKRVKDVKIGEEAKVADKEDIMSV"
283260. .284030
/gene="upps"
283260. .284030
/gene="upps"
/note="SAV1260"
/codon_start=1
/transl_table=11
/product="undecaprenyl pyrophosphatase synthetase"
/protein_id="BAB57422.1"
/db_xref="GI:14247030"
/translacion="MFKKLINKKNKNNYNEELDSSNIPHEIAIIMDGNRWAKKRRM
PRIKGHYEGQTKITRIASIDIGVYLTLYAFSTENSRPSEVNTIMLPVNFLLTK
FLPELIEKNVKVETIGFTDKLPKSTIEAINNAKEKTANTTGLKLIFAINYGGRAELVH
SIKNMDELHQOGLNSDIIDETVINNHLMTKDYPDELLIIRTSGEQIRISNFLINQVSY
SEFIENQKLPWDEDEBLIKICIYQSRQRRCGLSEE"
284037. .284819
/gene="cdsA"
284037. .284819
/gene="cdsA"
/note="SAV1261"
/codon_start=1
/transl_table=11
/product="phosphatidate cytidylyltransferase"
/protein_id="BAB57423.1"
/db_xref="GI:14247031"
/translacion="MKVRTLTAIIALIVFLPILLKGLGLVMPANILALIALKELLNM
NMKFEVSPGLISAVGLIIIMLPQOHAGPQVQIQKLSIAMSFEIVLSYTVLSKNRPF
MDAFCUMSVAYVIGIEFMFFETRESEGLHLYIAFLVILWLTDTGAYLFGKMMGRKHLW
PVSPNKTIEGFTGGLFCSLIVPLAMLYFVDNMMNWILLGVTLLLSLFGQLGDLVES
GFKRHFGVKDSGRLLPGHGGTLDRFDSFMFVLPILLNLIQS"
285031. .286317
/gene="SAV1262"
285031. .286317
/gene="SAV1262"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57424.1"
/db_xref="GI:14247032"
/translacion="MSYLVLTIIAIIIVFGLVIVVHEYGHMFFAKKAGIMCEPAIGMG
PKIFSFRKNLTLYIRLLPVGYVRMAGDGLPEPPVPPGMNKTIKLNEEIEITHILD
DHKFOQIEAIEVKCKDDLFIEGTAYDNERHFKIARKSPFVENGSLVQIAPRD
ROPAHKXPWKFLTFAGPLNFILALVLFGLAYYQGTPTSTVEQVADKYPAQOAGL
QKGDKIYQIGKYKISEFTDDVDKALDKVKNKTIVKPFERDGTKSVELTPKTKERKLTK
VSETKYVLGFQPASEHTLKPPIVYGESFLKGSTLIIFTAVVGMLASIFITGGFSFDM
NPGVGIYHNVDVVKAGIISLIGYTALLSVNLGIMNLIPIPALDGGRIIFYEIAIFR
KPVNKAETIIAIGAIFMVVIMILVTWNDIRRYFL"
286337. .288040
/gene="pros"
286337. .288040
/gene="pros"
/note="SAV1263"
/codon_start=1
/transl_table=11
/product="proline-trna ligase"
/protein_id="BAB57425.1"
/db_xref="GI:14247033"
/translacion="MKOSKVFIPTMRDVPSEAAQSHRLLKLSGLIKQSTSGIYSYLP
LATRLNNITAIIVQMERIDSVEILMPALQQAELWEESGRWAGYGPFLMRLODRHGR
OFALGPTHEELVTSIVRNELKSYQLPMTLFIQSKFERDEKRPFRGLRGREFIMKDA
YSFHADASLDQTYODMYQAYSRIFFERVGNARPVADSGAIGGSHTHEFMALSAIGE
DTIVYSKESDYAANIEKAEVVEPNHKSVPQLEK IETPNVKTQAOLADPLGRPYDE
IVTKMIFKQDGEYIMVLVRGHHEINDIKLSYFGTIDIELATQDEIYVNLVGNPGLG
PVIDKEIKIYADNFVDNNLVGNANEDGYHLINNVYGRDNFDVDEYDGFRILEGEKL
SDSGVAHFAEGIEVGQVFKLTGTYKYSMNATFLDNQKQAQITMGYIGIGISRTLSA

gene

299145..300116
/gene="ribC"
CDS
299145..300116
/gene="ribC"
/note="SAV1272"
/codon_start=1
/transl_table=11
/product="riboflavin kinase / FAD synthase"
/protein_id="BAB57434.1"
/db_xref="GI:14247042"

/translation="MKVLEVTHPIQSKQYITEDVAMAFGFFDGMHGHDKXVFDILNEI
AEARSLKAVMTPDPSPVVLNPKRRTTYLTPLSDKRIEKISQHDIDYICVNFSSRF
ANVSDFEVNIIKNKREVIAGDFDTGKFGKGMNTVLQEDAFNTTIVSQEEN
EKSTSIQDLINGELQANDALGIYISIKTVQGEKRGRTIGPTTANIQSDDYL
LPKGVAVISYEIGTENKLYRGVANIGVKPTFDHPNKAEVVIEVNIFFDPEDNIYGERV
TVNHHFLRPEIKFEDGIDPLVKQMNDKSRAYLLAVDFGDEVYNI"
300231..300500
/gene="rpsO"
CDS
300231..300500
/gene="rpsO"
/note="BS18
SAV1273"
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S15"
/protein_id="BAB57435.1"
/db_xref="GI:14247043"

gene

/translation="MAISOERKNEIIKEYRVHETDTGSPVOIATVLTAEINAVNEHLR
THKKDHSRGLLKMVGRHLLNLYRSKDIOIRRELIKSLGIRR"
300869..302965
/gene="pnpA"
CDS
300869..302965
/gene="pnpA"
/note="SAV1274"
/codon_start=1
/transl_table=11
/product="polyribonucleotide nucleotidyltransferase"
/protein_id="BAB57436.1"
/db_xref="GI:14247044"

gene

/translation="MSQEKVKFTWAGRLTITETQLAKOANGAVLRVGDVVLST
ATASKPRGDFPLVNVEEKMYAAGKIPGGKKREGRGDDATILARILRIPRL
FPGYKHVOIMNVLADPCSPQAMIGSSMALSDIPFGQPIAGVNVGVIIDGK
YIINPTVEKEVSRLEAVAGHKDANVMVAGASEITEQEMLEAIFFGHEEIQRLVDF
QQOIVHQPVKQETIPAERDEALVERVSKLTEEKGKETVLTDFKQORDENLDNKE
EIVNEFTDESDPENELLIKEVAILNELVKEEVRRIADEKIRPDGKKPDEIRPLDSE
VGILPTHSGSLPTRGTOALSVLTLGALGDYOLIDGLGPEERKRMHHVNFNFSVG
ETGPPAPGRREIGHGALGERALKYIIPDTADFPYIRIVSEVLENGSSQASIGCS
TLAMDAGVPKAPVAGIAGMLVTRDSYITLIDIQMEDALGDMDFKVAGTKEGITA
IOMDIIKIDGLTREIIEALEQARRGRLEIMNHMLQIDQPTLSAYAPKVVMTIKP
DKTRDIGFGKKINEIIDETGVKLDIEQDGTIFIGAVDAQMINRAREIIEETREAE
VGOTYQATVKRIEYKGFVGLPGKDALLHISQISKNRIEKVEDVLKIGDTIEVKITE
IDKGRVNASHRALEE"
303201..304874
/gene="SAV1275"
CDS
303201..304874
/gene="SAV1275"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57437.1"
/db_xref="GI:14247045"

gene

/translation="MSLRKKKKNDIRIIPLGVGGEIAKNMYIVEVDDEMFLDAGLMF
PEDEMIGDIIVIDISVLENNKDKLGIPLTHGHEHAIGAVSVVLEQLDAPVYGSKIT
IALIKNMKARNIDKKRVYTVNNDSIMRKPNVINISFFNTHTSIDPSLGCVHTSYGA
IYVTGEKFDOSLGHYAPDIIKMAELIGEVEVFLISDSEAEKPGSYNTPENVIEHHM
YDAFAKRVGLVSCYASNFIRIQVNLNASKNRKVSFLGRSSFNIAKRMGYFD
IPKDLIPTEVDNYPKNEVIIATIQMGQPEVLEASHQMAQHKHKINIEEGDSVFLLAI
TASANEVITIANLNEIYRAGAHIIIPNNKKIHASSHGMCEELKMNINMKPEYFIPQV
GEFKMIOIAHAKAAAEAGVAPEKIFLVEKGDVINYNGKMDILLNEKVSNGNLLIDIGIGI
DVCNIVLRDRLHLLAEDGIFIAVVTLPDKNRRIAAGPEIQSRGVYVYRESEDLLEAE
KVREIVEAGLQEKRIEWSKIQNNROOI SKLLFESTKRRPMIIPVISEI"
305131..307497
/gene="spoilIE"

CDS

305131..307497
/gene="spoilIE"
/note="SAV1276"
/codon_start=1
/transl_table=11
/product="sporulation-related protein SpoIIIE homologue"
/protein_id="BAB57438.1"
/db_xref="GI:14247046"
/translation="MAQAKKSTAKKKTASKKRTNSRKKNDNPIRVYIAILVVVLVW
LGVOYGLIGRLIDSPEYLFQYSRYLTYLVLATATGFTYSKRPTRTAGTSVLIQ
IALFVSQLVFHNSGIIKAREPVLVSYQSYQHSHPNFGGVLGYLLELSVPLIS
LFGVCIITILLCCSVILLNQHORDVAKVALENIKAWFGSNFKMSERNOEKOLKRE
EKARLEEQARQNEQPIKDVSDFTVEPQERDIPYIHTENESKSCQPSKRKRVED
KATSNVINHQADQQBLTEQTHNSVESNTIEEAGEVTNVSYVPLTLLNQPAKQ
AENSKAEVRQGOVLENTLDFGVNAKVTQIKIGPAVTEIOIQAQGVKYSKIVNLHN
DIALAALAKDVRLEAPIGPGRSANGIEVPNEKISLVLSKEVLDEKFPNNKLEVLGLRD
ISGDPITYPLNEMPHLLVAGSTGSGKVCINGIITSILLNAKPREVKMLIDPKMWEL
NYPNGIPHLPIVVPVNPDKAAQALEKIVAEEMERYDLFQHSSTNKGYNELIRKKQO
ELDEKQPELPYIIVIDEADLMWAGKEVENAIIQITOMARAAGIHLIVATQRPISVD
VITGIIKNIPISRIAFVSSQTDSTRTIIGTGAEEKLIGKGMLYVVGNGSDSOTRIQGA
FLSDQEVQDVVNVVVEQQQANYKMEPDPADVDKSEMKSEDALYDEAYLFWVEQOKAS
TSLIQBQFRIGYNRASRLMDLERNQVIGPQKSGKPRQVLIDLNNDEV"
307502..308215
/gene="SAV1277"
CDS
307502..308215
/gene="SAV1277"
/note="similar to transcription regulator GntR family"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57439.1"
/db_xref="GI:14247047"

gene

/translation="MSBMNAVNVKYOILNIMQNKLEYDQOLPSNLIAELNVKTD
DYVEAQALITEQVIKDNFEGTSVLSLPPEFVPLNELISIGOMIKKAGFECGTAYLN
FDEQATMLDANLLSVEEGYPVTLIERLRADGEPVYVCLDKAKKELTCTETQMSNG
SILSAIKQESNHNIYADTEIAVNEPRISEVLNASPHEGLILLKITHNEULDEPIL
YSLNYMKNLSLVQFKITRKI"
308246..309511
/gene="SAV1278"
CDS
308246..309511
/gene="SAV1278"
/note="similar to processing proteinase homolog"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57440.1"
/db_xref="GI:14247048"

gene

/translation="MSRQSQPNHIKVSPTTKTTTIVFKFMAPLEYDFTITARSLLS
KLIVRATKKWPTDKSFNNHLADLYGAVVNSTISKFKDQHVITFSLEIVNERYLRNGES
LFNQGLDLQEIWNPLIENKAFNDNFVNOEKTLAKKIEAMVDNKAQYSLKLLDHM
FNEQAYKLSGQLEQIPIHTAETLYHTYQSMINNDQCSVYVVGNVPEPSVEKQIREK
FALKPPDKHQFQSTHHLHDEVDYIYVEDDQAKLNMGYRPTQYOGSGYAAVYVF
FALKPGDPSVLPNEVREKOSLAYSHSQIDGRNGYLFVLVSGVSSDKYETAKDTIISE
NMFGDGFTEKEKLEAKVLIISHRYESEDPRKSIIEIMHNQILLEQPPQSKETFINDI
QKVSREDIVSAEKAFLDTIYVLTGGDK"
309511..310797
/gene="SAV1279"
CDS
309511..310797
/gene="SAV1279"
/note="similar to processing proteinase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57441.1"
/db_xref="GI:14247049"

gene

/translation="MKERYELIDERVFEQLENGLRLFIIPKPGFQKTVTVTTQPG
SLDNQKPLQDQGVTPDGVVAHFLEHLEKFEKEEDIFTAFEDNAQNAFTSFDRTS
YLFSAVDINNNIKRLLTWETVPTFKETVDKKGIIAEETKMYOEGQYKLMFNTLR
AMYQRHPIRDYIAGSVESIYDITKDDLYLCYETFYHPSNMVLFWGVNPEETCRIVK
QREDARKNVQPKIERGLVDPEDEKFAFTESMKIOSPRMLGFKNNKPLQEPAPYV
QHDENKSLFIFEGEETDYQNLLNEGLDDTGFYQVLEPTYYSFVTSVTSATEEPDK
LKLLDELRLKKGNFQDAEAFELLKKQFGEFISLNSPEYIANQTKLYLFEQVSFV
DMLDIVENITLDSINETSLSYLNLDQDQDSRLKIKK"

gene
CDS

```
310797. .311501
/gene="SAV1280"
CDS
310797. .311501
/gene="SAV1280"
/translation="similar to 3-oxoacyl- acyl-carrier protein
reductase homolog ymfI"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57442.1"
/db_xref="GI:14247050"
/translation="MKALVGGSGSIGSEIVKQLLTGDFEVYVYYRTDINELTSKFN
DDKVRFTQADLSQITIDKTFGDKSLDCLYASGSLYGLQDKKDDHDACVQLNV
LQLRLCRYFVDVLRQANGRIIYISSIGETGASMETIYSAMKSAQLGFKALSQEL
ALTSVTYNATAPGVFVAGNMASEWOEQLQAMITELPQORLVLPSEVAHTCAVLYHSNA
RSVTGTIQKNGAWYI"
311606. .312433
/gene="SAV1281"
311606. .312433
/gene="SAV1281"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57443.1"
/db_xref="GI:14247051"
/translation="MTVAEKKEMWLEYEIEINRPLLGDVSSLLGMLGISIVTINGVD
QGRGLIKTDNLEKVERFEQIARGINEIETIKLKPDLRDLAVRHGRVTEQDAKDK
KTFPEREDLGLVDFLAELFKEGHKLIGIRGMPRVGKTESI VAGSVCAHKKRWLFIS
STLIKQVRSLLKGEVDANHVYIIDGAVTARSNPKHOELVNEVMTLPSIKVVEHPD
LFVETSTCTWEDFYIILRENEHQEIHYEEMKKQTQVSRNNLDFGPPFGGGFFFE"
312452. .312844
/gene="SAV1282"
312452. .312844
/gene="SAV1282"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57444.1"
/db_xref="GI:14247052"
/translation="MKTVEGALKGRRLRGLMTLEBQRTGIKREMLVHTENNEFDOL
PNKNYSGFIRKVASVNIIPNOLIOAHQEIIPSNQAEWDEVITVPNNKDLQYKSKS
KEPIQLLVINGITVLTILLWIMLVLP"
312878. .313456
/gene="pgsA"
312878. .313456
/gene="pgsA"
/translation="MNIPNOITVPRVVLIPVFILFALVDFGNVSFLGCGYEIRIELL
ISGFIFILASLDFVDGYLARKNNLYTNMGKFLDPLADKLLVASALIVLQGLTNSV
VAIIITAREFAVIGRLQLIQEGFVSAAGOLGKTKTAVTWAITWLLGLDPLATLIGL
SLQOILYIGVITFIIUSGIEFYKGRDVFQKQ"
313680. .314831
/gene="cinA"
313680. .314831
/gene="cinA"
/translation="MSIAITAVGSELLGQIANTNGQFLSKVFNETGQNVLENHKVIGD
NKARLESSVRHALEKQVILVTIGLGLPTKDDLTAKHTVAQIVGKDLVIDEPLKIESY
FEPQGGEMTPNNKQALVIEGSTVLANHGMGAPCMYFNENKQIILLPGPPKEMQPMV
KNELSHFINNRRIIHSELLRFAGIGESKVETIILIDIKQTNPTIAPLAGSHEVIR
LTNADSKEOAQLIOPKVEILDRIGEYVYGSDDTLIEQAVIKKTHEPVIYDGTIN
GALYHRLKEVDLNLKGMINHNENFVINKPLEQOLKDAQOVNKLNVSSAIIILE
YDGVHLIGYDNNEFFTEQFKMSKRNLLKNRSQNVILIRLLNWLRTIN"
```

gene
CDS

```
314996. .316039
/gene="reca"
314996. .316039
/gene="reca"
/translation="MNDROKALDVTIKNMEKSGKGAVMKLGDNIGRVSTTSTGVS
TLDNALGVGGPYKGRITIEIYGPSSGKTTVALHAIAEVQSGNGVAAPIDAASHALDPEY
AQALGVDDINLIYQSDPHGEGGLEIAEAFYRSGAVDIVVDSVAALTPKAEIEGEMGD
THVGLQARLSQALRKLKLSGATSKNTTAIFINQIREKGVNMGFPETTPGGRALKFYS
SVRLVERAEQLKQGEIVGNRTIKIVVKNKVAPPFVRAEVDIMYGOCISKEGELIDL
GVENDIVDKSGAWSYNGERMGQCKENVMYLNKPNQIKBEIDRKLREKLIGSDGDVE
ETEDAPKSLDDEE"
316393. .317952
/gene="SAV1286"
316393. .317952
/gene="SAV1286"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57447.1"
/db_xref="GI:14247055"
/translation="MNDROKALDVTIKNMEKSGKGAVMKLGDNIGRVSTTSTGVS
TLDNALGVGGPYKGRITIEIYGPSSGKTTVALHAIAEVQSGNGVAAPIDAASHALDPEY
AQALGVDDINLIYQSDPHGEGGLEIAEAFYRSGAVDIVVDSVAALTPKAEIEGEMGD
THVGLQARLSQALRKLKLSGATSKNTTAIFINQIREKGVNMGFPETTPGGRALKFYS
SVRLVERAEQLKQGEIVGNRTIKIVVKNKVAPPFVRAEVDIMYGOCISKEGELIDL
GVENDIVDKSGAWSYNGERMGQCKENVMYLNKPNQIKBEIDRKLREKLIGSDGDVE
ETEDAPKSLDDEE"
316393. .317952
/gene="SAV1286"
316393. .317952
/gene="SAV1286"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57448.1"
/db_xref="GI:14247056"
/translation="MNLISLLILLGTLGVVGVVVARNLLLOKQSOAROTAEIDVN
OAKKEADNIKEKLEKEKEOILREOTEAELRERSELOQETRLLOKEENLERKSD
LLDKKDILOKESKIEEKQOQVDKAKESSVOTLIMKHEQELERISGLTQEERINEQLQ
RVEEELSQDIAVLVKEKEKEKAKEDIVKTAKELATAVORLAADHTSESTVSVNLPND
EMKRIITGREGNRTIETLTGIDLIIDDTPEAVILSGFDP IRREIARTALNVLSQD
RIHPGRIEDMVERAKREVDITIREAGQATFEVNAHNMHPDLVKIVGRLNRYTSYQON
VLKHSIEVALHSLMAAEELGEDETAKRAGLLHDVGRAIDHEVEGSHVEIGVELAKKY
GENETVINAHSHHGQDVETPTIISILVAADALSAARPGARKETLENYIRLERLETL
SESYDGVEXAFATQAGREIRVIVSPEDDLKSYRLARDIKNQIIDELELQTPGHKVTV
VRETRAVEYAK"
complement(318249. .318419)
/gene="SAV1287"
complement(318249. .318419)
/gene="SAV1287"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57449.1"
/db_xref="GI:14247057"
/translation="MSNPSKFEDAEDIAKESLKLYINAYRLAERYSTIATNCYDTAEV
IKELKTDLOLK"
318638. .319435
/gene="SAV1288"
318638. .319435
/gene="SAV1288"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57450.1"
/db_xref="GI:14247058"
/translation="MRIMFIDGIVGKIGRDAIETYIPQLKOKYPTVTIVNAENAHG
KGLTEKIYKOLLRNGVDMFTMGNTYQRELYDFIDEAKRLVRPANPDEAPGIGMRP
IQINDIKLAVINLQGRAPFMDIDDPFKADQLYKEAQETPFIVDFHAETTSKYAM
GWLHDGRASAVGTHTHIQIADERILPKGTGITYDVGMTGYDGLIGINKINTEVERFI
TSLPQRHVPNEGRSVLSGVVIDLKEGKTKHIERILINDDDHPFSTF"
319575. .321335
/gene="SAV1289"
319575. .321335
/gene="SAV1289"
/translation="similar to 2-oxoacid ferredoxin oxidoreductase
alpha subunit"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57451.1"
/db_xref="GI:14247059"
/translation="MKPQLSNKVGQGGQGGEGTGTATAMNKKGYLYLYGRHFSSR
```


IKGHTNNKIRVSTTPVHAISDDLDLIIAFDQETIDVNHHEMREDSIIILADAKAKPVK
PECHQAQLIELPTTAKELGTALMKNMVAIGATSAIMNLNLTNTFEELIYNMFSSKKGD
KVVEVNIQALNEGYQLMQSLRPEIDGDFELEDSTDALPHLYMIGNDAIGLGAIAAGSQF
MAAYPTIPASEVMEYMIANISKVNGAVIOIPEDEIAAVTMAIGANYGVRAFTASAGPG
LSLMAAIGLSCMTETPLVINIQRGSPSTGLPTKQEQSDLMOMIYTHGDIPIKVIYA
PTDAEDAFITLMEAFNLAEQYQCPVIVLSDLQSLGQKTEKLDYNRIEIKRGEIIQS
DIEREDDDQFKRYALTNGVSPRIPGVKGGIHHITGVHEHNEGKPSSESANROQQ
MEKMRKIIEOLLIESPEANLQHEADAILYIGFISTKGAIQEGSNRLNQOGIKVNTIQ
IRQLHPETSVIQODAVNAKAKVVVVEHNYQOQLASIIKMNVNTHDKIENTYKYDGTFF
LPHEIBEKGKIIATEIKEMY"
321335..322202
/gene="SAV1290"
321336..322202
/gene="SAV1290"
/note="similar to 2-oxoacid ferredoxin oxidoreductase beta
subunit"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57452.1"
/db_xref="GI:14247060"
/translation="MATFKDFRNNVKNMPCGCGDFSVQAAIOKAAANIGLEPEEVAI
ITGICSGRSLGYINSYGVHSIHGRALPLAOGYKMAKDLTVIASGGDGDGYAIGMGH
THALRRNNMITYIVNDNQIYGLTKGOTSPSSAVGFVTKTTPRKNIEKNVAPLELALS
SGATFVAQGFSSDIKGLTLIEDAIINHDFGSFVNVSFPCVTYINKINTYDMFKELHLSV
DDIENDYSDTKQLATKTVIEHESLVTGIYVQDKTPTSYESQIKELDTPPLAKRDIKT
EDTFNALTEQFI"
322296..322589
/gene="SAV1291"
322296..322589
/gene="SAV1291"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57453.1"
/db_xref="GI:14247061"
/translation="MKDTLMSIQIIPKTPNNDNVIPYVDEPAIKIIDESGLHFRVGPLE
TVVQGNMNECLILIQSLNRMVELECPISIISQVKFYHVPDGIPIETLTEKYDE"
322723..324267
/gene="SAV1292"
322723..324267
/gene="SAV1292"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57454.1"
/db_xref="GI:14247062"
/translation="MNEEQKASSVDVLAERDKKAEKDYSKYFEHVYQPPNLEAKKR
GKOEYRNRDFQIDERYGMGNERTELIRTYGQMNAHDETVIAGILEALGYQATTDI
NTADVILINTCAIRENAENKVFSEIGNLKHLPKRPDIILITGVCMSQSBESVYNNKILK
SYONVDMIFGTYHNIHHLPEILEAYLSKAMVVEVSKEGDVIENLPKVRGNIKAWVN
IMYGCDKFCYTCIVPTRGKERSRPEDIIDEVRELARSGYKEITILGQNVNSYKDL
ODIYDLGLLQALISAIPRVRTTSHFWDFDHMDIVISEGNNIVPHLHPVSGN
NAVILTMGRKYITRESYLDLVKRIKRIPNVALPTDIIVGYPNESEQIAPETLFLYDEV
GFPHATYIYSQRDQTPAAKMDKNVPLKMRQRLNKKVGHYSQIAKMSYKEGQVTV
VLCEGSSKKDDQVLAGTYTDKNKLNVFKAPKEMIKGVLEVRIDEAKOYSLNGSPVKEVE
PENVIQ"
324268..324633
/gene="SAV1293"
324268..324633
/gene="SAV1293"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57455.1"
/db_xref="GI:14247063"
/translation="MYNKDDVLKQADNIAINKNLNDITIKTYQOIEAQIHQNOTIKTKM
DMLKKHQKQAVNFQYQKONALEQSEHTIQSTEAENTLPIVEQFQTSQYEANQLLKM
FVSTMETRLNDHNKAKHSD"
324660..325151
/gene="SAV1294"
324660..325151
/gene="SAV1294"

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57456.1"
/db_xref="GI:14247064"
/translation="MKSRLKAITALLIAINVLSSIIIIPLGPVKAAAPVQHVFVNVLSA
VIVGPNYGAQALISSITRLVFGTGFAPAFPQSMIGVLLASMFYIYRKHFHMAAVGV
LGTGVIGSLICIPLAYFLGQDFFIKPLMITFVSSAIGSIISYFLLITLTKRGILQR
FIK"
325454..328073
/gene="mutS"
325454..325756
/gene="mutS"
/note="SAV1295"
/codon_start=1
/transl_table=11
/product="DNA mismatch repair protein"
/protein_id="BAB57457.1"
/db_xref="GI:14247065"
/translation="MSNVTPMQOYLKIKSEYQDCLIFRLGDFYEMFYEDAKEASRV
LEITLTKRDAKKENPIPMCGVPYHSADSYIDTLVNNGYKVAICEQMERSETNERYG"
325749..328073
/gene="mutS"
/note="SAV1296"
/codon_start=1
/transl_table=11
/product="DNA mismatch repair protein"
/protein_id="BAB57458.1"
/db_xref="GI:14247066"
/translation="MYRREVVRIVPTGTVMEQGVDDQKNNYILSFVMNQPEIALSYC
DVSTGELKVTHFNDEATLLNETIINPNEVYINDNISDNLKRQINMYETITVRETLS
SEIYSVNQOTEHLMQATQLLDYIHHTQKRDLSHIEDVYVYAAIDYMKMDFYAKRNL
ELTESIRLSKKCTGLLWMDETKTPMGARLKWIDRPLISKEQIPEARLDIVDFEFSAH
FIERDLTRYLVNQYDIERLYGVRSYGVNARDLIQLKHSISEIPNPKALLNSMNQNT
LVQVNOLEDDLDLILEQSVKADGGFLKVGFTSEIQLSEIPNLEASKNKGTWLA
ELQAKERQTGIKSLISFNKVFYFIETIRANLQNFPESEFGYMRQTLNNAERFTIT
DEKKEKEDIILGAEDKAIIELEYLFQVREEVKQYTERLQQAQAKIISELDCIASFAEI
AOKYNTTRPSFSENKTELVESRHPVVERVYNDYVPPNCRDNDETFFIYLTGPNNKS
GKSTYMRQVAIIISIMAQMGAYVPCKEAVLPIDQIFTRICAADDLVSXGKSTFVMELE
AQKALTYANEDSLLIIFDEIGRGTSTYDGLAQAAMEYVAETSHAKTLFSTHHELT
LDQALPSLKNVHAANEYKGLIFLHKVKGADVDDSTGIGVAKLADLPKVISRAQVI
LSEFEASAGKKSSISNLKMWNEPEINQENLISVBEETDTLSQKDFEQASFDLPFEND
QSESEIQLIKNLNSMNTPIEALVKLSELQNLK"
328086..330095
/gene="mutL"
328086..330095
/gene="mutL"
/note="SAV1297"
/codon_start=1
/transl_table=11
/product="DNA mismatch repair protein"
/protein_id="BAB57459.1"
/db_xref="GI:14247067"
/translation="MGKIKELQTSLANKIAAGEVERPSSVVKELLENAIDAGATEIS
IEVEESGVOSIRVYDNGSGIEAEDGLVFRHATSKLDDQEDLFHIRTLLGRGEALAS
ISSVAKVTKCTDNANGNEIYVENGEILNHPAKAKGDTDLVESLFYNTPARLKYI
KSLYTELGKTDIVNRWAMSHPDIRIALISDGKTMLSGTNGSGRTNEMVAEIIYGMKVAR
DLVHISGDSYDHIEGFVAKPEHSRNKHYISIFINGRYIKNPMNLKAILLEGVHTLLT
IGRPIIINIEMDPIILVDVNVHPTKLEVRLSKEEQYLIVSKIQEAFKDRILILPKN
NLDYVPKKNKLYSEFQKTEFEQRONTNNEQKTFSSSEESKSPMAENQNDIEVIK
EDSYNPFVTKTSLSLITDDDESGYNTNREKDEDYFKKQOEILQEMQDTFQSDNEDASVQ
YENKASDDYDVNDIKGTSKDPKRRIPYMEIVGVGHGYIIAQNEFGMYMIDQHA
QEKIKYEYERDKIGEYTNVEQDILLPLTFHESKDEOLVIDOYKNELOQVGMILEHFGG
HDYIVSSYPVWPKPEVEEIIKDMIELLIEEKVDDIKKREDVAMMSCKSKSIKANHY
LQKHMSDLIDQUREADDPETCPHGRPIIINFYSKYLELEKLFKRVN"
330110..330643
/gene="g1pp"
330110..330643
/gene="g1pp"
/note="SAV1298"
/codon_start=1
/transl_table=11
/product="glycerol uptake operon antiterminator regulatory

```
protein"
/db_xref="GI:14247068"
/translation="MNNILPAIRNIKIDLEKLIKTDYKMCVLLDMHGHIGHSIMELK
QNHTECFIHIDLIKGLSHDEFSEFIIOQYKPGIVSTKSKVIAKKASLNTLTIFRVE
IIDSQALKRSDLDLJKVPEPDEVLPGVASKAIHHIQNETNTQVIAGLNTIDEVNE
AVKNGAKVYVTSYDKLW"
complement(330680..330769)
/gene="SAV1299"
complement(330680..330769)
/gene="SAV1299"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57461.1"
/db_xref="GI:14247069"
/translation="MLENYSLYDFTLNTLLHIPYENGFFIMLYD"
331118..331936
/gene="glpP"
331118..331936
/gene="glpP"
/note="SAV1300"
/codon_start=1
/transl_table=11
/product="glycerol uptake facilitator"
/protein_id="BAB57462.1"
/db_xref="GI:14247070"
/translation="MNVYLAELFTAILILFGGVCANVNLKRSAAANGADWIVITAGW
GLAVTMGVAVGFGSAHLNPAVSLALGSPDMSLVPGVIVAOMLGAIVGATIVWL
MYLPHWKATEAGAKLQGVSTAPAIKNYFANLSEIIGTMALTLGILFIGNKTIADGL
NPLIVGALIVAGLSLGGATGYAINPARDLGPRIAHAILPIAGKGSNWSIAIVPIIG
PIAGLGLGAVYAVFYKHTFNIGCAIAIVVVIITLILGYILNKSRRKGDIESY"
332065..333561
/gene="glpK"
332065..333561
/gene="glpK"
/note="SAV1301"
/codon_start=1
/transl_table=11
/product="glycerol kinase"
/protein_id="BAB57463.1"
/db_xref="GI:14247071"
/translation="MEKYILSIDOQTSSRAILFNKGELTAGVAOREFKOYFQSGHV
EHDANEIWTSLVAVMTVEVINENVRADQIAGIGITNORETIVVMDKHTGPRTHAIVW
QSRTOISSELKQOQYEFTDRKTLGLLDPYFAGTKVKWILNDVEGAREKANGDLL
FGITDVLWVWLSGKAHITDYSNASRTLMFNHIDLEWDDLELLTVPKMLPEVPK
SSEIYGTIDYHFYQGVPIAGVAGDQQAALFGQACPECGDVKNYGTGTGFMMLNTGD
KAVKSEGLLTTIAYGIDGKVNVALEGSIFVSGSAIOWLRDLGRMINSAPOQSEYATR
VDSPEGVVPVPAFVGLGTPYWDSEARCAIFGLRGTKEHEFIRATLESICQTRDWE
AMSKDSIDVQSLRVLDGGAVNNFMQFQADIVNVSVERPEIQETTLALGAFLAGLAV
GFWESKDDIAKNWKLKLEKFPDKMDEGEREKLYRGWKKAVEATQVFKTE"
333671..335392
/gene="glpD"
333671..335392
/gene="glpD"
/note="SAV1302"
/codon_start=1
/transl_table=11
/product="aerobic glycerol-3-phosphate dehydrogenase"
/protein_id="BAB57464.1"
/db_xref="GI:14247072"
/translation="MWDKSLVFTFFRRFRGMALSTFKREHKKNLRNDEYDLVIIGGG
ITGAGIALDASERGMKVALVEMQDFAQGTSSRTSLKLVHGLRLYKQFQIGVAETGKE
RAIYENGPHVTPPEWMLPMHKGFTGKFTSIGLGMYDLRAGVKKSERKMSLRSKE
TLAKEPLVKGELKGGYVVEYRTDDARLTIEVMKRAAEKGAETINVTKSEHTYDKN
QOVNGVVIDKLTNENVTIRAKKVNAAGPWDDVRSDDYARNKKLRLTKGVHVVID
QSKPELGOAVYEDTEKDGRIIFAIPREGKAYVGTDFPDYDNIKSSPLTTQEDRDYLD
AINYFFSVNWDIESTWAGIRPLIYEGKQFSEISRKDEIWEKSGLLTITAGGL
TGVYHMAQDILDLVLSRKLKXDYGLTFSPCNTKGLAISGGVGGSKNPDFAFVEQKVDVA
KGFGIDEDVAVLSKSGSNVDELNFIAQTSQYHDSKLPLEIYVLYSITQCEWYKP
NDLVLRRSGKWFYNIKDVLQYDKAVIDIMADMLDYSPAQIEAVTEVEEQAIKEAQHGN
NOPAVKE"
335542..336456
gene
```

```

CDS
/gene="SAV1303"
335542..336456
/gene="SAV1303"
/note="similar to lysophospholipase"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57465.1"
/db_xref="GI:14247073"
/translation="MTEKQFKLTVDQNTNTEVKVNFDTVDVSKGLIHIFPHGMAEHMERY
DKLAHALSKHGFDVIRHNHRCHGINIDESRGHVDDMKRVIDGAFEVAQTVRGVNDKP
YIILHSGSVIARLFLVETYPQYVDGLLSTGMYSLWKGLPYTKVQLQITKIYGAEK
RVENWNLVSNENKKIRPLRTQSDWISSNPIDSSNPIEDNFIDKPYSGNFNQLLIYQTAIY
MLHTSQLKNMKMLNHAMPILLVSGYDSDSLGDYSGKILKLANIYRKAGIKNKVKNLYHH
KRHEVLFEKDHDKIWEDLFKWLNFYKK"
336474..337409
/gene="miaA"
336474..337409
/gene="miaA"
/note="SAV1304"
/codon_start=1
/transl_table=11
/product="tRNA delta(2)-isopentenylpyrophosphate
transferase"
/protein_id="BAB57466.1"
/db_xref="GI:14247074"
/translation="MNNKPKPIVIVVGTASGKTELSIELAKRINGEIIISGDSMQVYK
HMNIGTAKVPPEEMDGPHLIDILNPDPTFSAYEFKRLAEDLITDITNRGKVPITAG
GTGLYIGSLYVLEDEDETVPQAOLSVKOKLSALEHLDNQQLHDYLAQDVAESAENI
HPNRQVRLAIEYLYKTKLLSNRKKVQOFTENYDTLLIGIENSRTKLSRINKRVD
IMDHGLFNRKQVLEQGYECSQMAIGIKELIPVINGOMIYEDAVDNLKQHSRQYA
KRMQTDLFREKMSVHLDKKNNSLQMLDEITTIQK"
337424..337657
/gene="SAV1305"
337424..337657
/gene="SAV1305"
/note="similar to host factor-1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57467.1"
/db_xref="GI:14247075"
/translation="MIANENIQDALENFANQKQTEVTVFVFLNGFQMKGVIEEDKYVY
SLNSQKQHLIYKHAISTYVTEGQASTESE"
complement(337878..338354)
/gene="bsaA"
complement(337878..338354)
/note="bsaA"
/gene="bsaA"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase"
/protein_id="BAB57468.1"
/db_xref="GI:14247076"
/translation="METIYDFVETNKGVTYKLDAYKGDVMLIVNTASEGFTSQFEG
IQSLYERKDGFEVLGFCNCFGQEPGSEEAQNCNLNYGVTFPMHQKIDVKGGH
QPLPFRYLTAOHGFFNEKIKWNFTKFLVDREGNVVRRFAPQKRPQIERIEKLL"
338466..339704
/gene="SAV1307"
338466..339704
/gene="SAV1307"
/note="similar to GTP-binding protein proteinase modulator
homolog yrbA"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57469.1"
/db_xref="GI:14247077"
/translation="MAQQQIHDTKNKLKAVLVGVHAQDDKQFNFEEMBELSSLSST
COLVGLQIQNRDRDVRKYVVGKITEETQAFTEFDIDVITNDLTTAQSKSLNE
ALGVKIIDRTQILILEIPALRARSKEKLOVELAQLDYLLPLQCHGKSLRGGIGCT
RGPETKLEMDRRHIRTMRNEIKHOLTVEHREYRNKRNOQVFOVALVGYTNACK
SSNFVLANEETYEKQDLFATLPKTRQIQINDGFNLIISDTVGFQIKLPTTLIAFK
```

```
STLEAKGADLLVHVVDSSHPYRTOYDTVNDLIKQLDMSHISQIVIFNKKDLCDHAS
NRPASDLNPNFVSSKNGDKLVLKTLFIDEIKRQLTYTYDETATNADRLYFLKQHTL
VPELYDEIENVYRIKGEKK"
339723..340961
/gene="SAV1308"
CDS
339723..340961
/note="similar to aluminum resistance protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB5470.1"
/db_xref="GI:14247078"
/translation="MKDISKIIVADVESTLAPYFKEIETAYINQKVLNAPHHVKATE
SDLOSTGYDGFGRDHEIYAQAFKAEDAIVRPQIIISGTHAIIAQSLLRKHGDE
LIYITGSPVDTLLEIVGNGIESLMHGVSYKDIALKEGKIDIESVLGDGVSRTKV
TAIORSKYDORPSIPLDEIEKVIITLKNVHPNILLIFVDCVGEVEREPIECGADL
TAGSLIKPFGGLAKIGGIAGRKOLIERCGYRLTAPGIGKEAGASLNALLEMYGFF
LAPHVVSQSLKGAFTSLFLEKMNMTTPKYIEKRTDLIQTVKFKQOMISFCOSIQ
HASPINAHSPEPSYMPGYEDDVMAAGTFIQGSSIELSADGPIRPPYEAIVQGGITY
EHVKITAVTRAVNQLKEQGLI"
341204..341572
/gene="glnR"
341204..341572
/gene="glnR"
/note="SAV1309"
/codon_start=1
/transl_table=11
/product="glutamine synthetase repressor"
/protein_id="BAB5471.1"
/db_xref="GI:14247079"
/translation="MISNDAIRNMAVFSMSVSKLTDLTPRQIRYETHELIKPERT
EGOKRLFSNLDERLLEIKSLLEKGFNIKGIKIIYDSQEHLLTTDEQEIIRKKMIVDAT
OKPIGETLPINRGDLRFIK"
341591..342931
/gene="glnA"
341591..342931
/gene="glnA"
/note="SAV1310"
/codon_start=1
/transl_table=11
/product="glutamine-ammonia ligase"
/protein_id="BAB5472.1"
/db_xref="GI:14247080"
/translation="MPKRTFTKIRKAEENVRVLRLOFTDILGTIKNVEVPVSQL
EKVLNEMWFQSSIEGEVRIEESDMYLPDLDTWVIFPWTAGQKVARLICDVKYTD
GTPFGDGRANLRVKLMEDLGFTDFNLGPEPEFFLFKLDKEGPTLELNDGGYFD
LAPTDLGENCRDVILELEDMGFDIEASHHEVAPQOHEIDFKYADAVTACDNIOFTFL
VYKTIARKHNHATMPKPLFGVNGSMHFNVSLEFKGENAFDPNTEMGITETAYQF
TAGVLKNARGFTAVCNPLVNSYKRLVPGYEAPCYIAWSGKNRSLPIRVPSSRGLSTRI
EVRSDPAANPYMALAAIILEAGLDGINKLKVPEPVNONIYEMNREEREAVGIQDLP
SLYALKAMRENEVIKALGNHIYNQFINSKSIENDYIRTQVSENERDOYMKQY"
344826..345032
/gene="SAV1311"
344826..345032
/gene="SAV1311"
/note="SAV1311"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB5473.1"
/db_xref="GI:14247081"
/translation="MNEIETIISEIEKLITNNTPYISKNSGVPRQTVTDLKVGTKI
KEAKFKTIKLYEORTLENKTEC"
346266..346451
/gene="SAV1312"
346266..346451
/gene="SAV1312"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB5474.1"
/db_xref="GI:14247082"
/translation="MSEYKKKIIEILIESNLTYEISKKTGVSQYVLSQLRQGRKREVDN
```

```
ITLNTTEKLYEYANKVL"
BASE COUNT 124661 a 48180 c 63742 g 110652 t
ORIGIN
AP003361 Length: 347235 February 15, 2002 13:42 Type: N Check: 3057
Found using 'seq4' (pappu403.key)
...
31637 ACATCAAAAAAGCAGACATAAGCAAAATACATAAAGATTCAATTGATAGCAGCAACGTTT
31687 31694
31697 AAAAAATAGTTTATCATCGCATTTTAGAACAGAGAAACCGTGATGTTTAAT
...
118299 GTCAACCTAAATTAATTTTAAATTTTTTTTAAAAAGTATGAATTAATATGTTGCAACGTTTCGAT
118349 118356
118359 TTAATATTAAAGAAAAAGAGACAGCAAAATCATGATGTGATTACGTGCT
...
128767 TGGTCAAGCGAGACGATTTTACATTAAAGATGAAAAATGTCATACAAAATGCCGAACGTTAA
128817 128824
128827 GCCAAAAGCAGATAAGAGTAATCCAAAGTCATATCGGTGCTCAAAATGCC
...
149942 CCAAGTTGATATCGGAGAGAACGACCTGTACAAAATCGTATGTGGTGCACCGAACGTTGA
149992 149999
150002 TGCAGGACAATATGTCATTGTTGCTAAAGTAGGTGGCAGATTGCCCTGG
...
239039 GGTGGAATAGTGAATCAATTAAGTGGGTATATCAAGTAGACGTTAATGCCGAACGTTTC
239089 239096
239099 AATACAAAACCAGCAGGATTATTAGAAAAGAAAAATTTTACCCTGGTA
...
340163 TGAAAGTGTTTAGATGGGTTTCTGAGCGCACCAAGTAATAGCGATTCAACGTTTCGAA
340213 340220
340223 AGGCTATGATCAAGACCTTCAATCCGCTAGATGAATTGAAAGGT
...
1 match found in sequence:
ar146299 ; TOIG of: ar146299 check: 5318 from: 1 to: 20
(from "immun_ge.seq")
TOIG of: ar146299 check: 5318 from: 1 to: 20
LOCUS AR146299 20 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 11 from patent US 6218371.
ACCESSION AR146299
VERSION AR146299.1 GI:15109488
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20)
AUTHORS Krieg,A.M. and Weiner,G.
```

TITLE Methods and products for stimulating the immune system using immunotherapeutic oligonucleotides and cytokines
JOURNAL Patent: US 6218371-A 11-17-APR-2001;
FEATURES Location/Qualifiers
source
1..20
BASE COUNT 4 a 7 c 3 g 6 t
ORIGIN

AR146299 Length: 20 February 15, 2002 13:27 Type: N Check: 5318 ..
Found using 'seq4' (pappu403.key)

1 ATCGACTCTCGAAGCTTCTC
10 17
|-----|
|-----|

1 match found in sequence:
ar148617 ; TOIG of: ar148617 check: 7567 from: 1 to: 14
(from "immun_ge.seq")
TOIG of: ar148617 check: 7567 from: 1 to: 14

LOCUS AR148617 14 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 11 from patent US 6225292.
ACCESSION AR148617
VERSION AR148617.1 GI:15112707
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 14)
AUTHORS Raz, E. and Roman, M.
TITLE Inhibitors of DNA immunostimulatory sequence activity
JOURNAL Patent: US 6225292-A 11-01-MAY-2001;
FEATURES Location/Qualifiers
source
1..14
BASE COUNT 4 a 4 c 2 g 4 t
ORIGIN

AR148617 Length: 14 February 15, 2002 13:28 Type: N Check: 7567 ..
Found using 'seq4' (pappu403.key)

1 AATTCAACGTTCCG
6 13
|-----|
|-----|

2 matches found in sequence:
ax076142 ; TOIG of: ax076142 check: 5564 from: 1 to: 6003
(from "immun_ge.seq")
TOIG of: ax076142 check: 5564 from: 1 to: 6003

LOCUS AX076142 6003 bp DNA PAT 06-FEB-2001
DEFINITION Sequence 9 from Patent WO0104318.
ACCESSION AX076142
VERSION AX076142.1 GI:12710795
KEYWORDS
SOURCE Myxoma virus.
ORGANISM Myxoma virus
Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
Leporipoxvirus.
REFERENCE 1 (bases 1 to 6003)
AUTHORS Mcfadden, G.
TITLE Myxoma virus genes for immune modulation
JOURNAL Patent: WO 0104318-A 9 18-JAN-2001;
Viron Therapeutics, Inc. (CA)
FEATURES Location/Qualifiers
source
1..6003
/organism="Myxoma virus"
/db_xref="taxon:10273"

BASE COUNT 1815 a 1469 c 1403 g 1316 t
ORIGIN

AX076142 Length: 6003 February 15, 2002 13:31 Type: N Check: 5564 ..
Found using 'seq4' (pappu403.key)

1 ATGTATCGAACGTTATTATTCTAATAATGATAGGATACGGCTATCCCGGAGGAAGTCGT
7 14
|-----|
|-----|

61 TACA
...
|-----|
|-----|

1737 AACCGCTCCTTCTACGACCGTAACCGTCGACACTAGTAGCGGCGTGTACGCGAACGTTCT
1787 1794
|-----|
|-----|

1797 GAGAACTCCTAAAGATGTGGAGGTGACCGCTCCGAAAGACGTCACCGT
...
|-----|
|-----|

1 match found in sequence:
ax076147 ; TOIG of: ax076147 check: 3310 from: 1 to: 903
(from "immun_ge.seq")
TOIG of: ax076147 check: 3310 from: 1 to: 903

LOCUS AX076147 903 bp DNA PAT 06-FEB-2001
DEFINITION Sequence 14 from Patent WO0104318.
ACCESSION AX076147
VERSION AX076147.1 GI:12710800
KEYWORDS
SOURCE Myxoma virus.
ORGANISM Myxoma virus
Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
Leporipoxvirus.
REFERENCE 1 (bases 1 to 903)
AUTHORS Mcfadden, G.
TITLE Myxoma virus genes for immune modulation
JOURNAL Patent: WO 0104318-A 14 18-JAN-2001;
Viron Therapeutics, Inc. (CA)
FEATURES Location/Qualifiers
source
1..903
/organism="Myxoma virus"
/db_xref="taxon:10273"

BASE COUNT 252 a 195 c 206 g 250 t
ORIGIN

AX076147 Length: 903 February 15, 2002 13:31 Type: N Check: 3310 ..
Found using 'seq4' (pappu403.key)

33 GTCGTTGTTACAGTCGTATGTACTAGCCGATGTAAGAACGATTTTAAGTCGGAACGTTCA
83 90
|-----|
|-----|

93 AGCTGTGAATAAGAAACAGACGCTACAACAAGACGAGGTAACGGAAT
...
|-----|
|-----|

1 match found in sequence:
ax083675 ; TOIG of: ax083675 check: 8143 from: 1 to: 22
(from "immun_ge.seq")
TOIG of: ax083675 check: 8143 from: 1 to: 22

LOCUS AX083675 22 bp DNA PAT 28-FEB-2001
DEFINITION Sequence 1 from Patent WO0112223.
ACCESSION AX083675
VERSION AX083675.1 GI:13185407
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.

```
REFERENCE 1 (bases 1 to 22)
AUTHORS   van Nest,G.
TITLE     Methods of modulating an immune response using immunostimulatory s
          equences and compositions for use therein
JOURNAL   Patent: WO 0112223-A 1 22-FEB-2001;
          Dynavax Technologies Corporation (US)
FEATURES  Location/Qualifiers
          source 1..22
              /organism="synthetic construct"
              /db_xref="taxon:32630"
              /note="Synthetic construct"
BASE COUNT 6 a 3 c 7 g 6 t
ORIGIN
AX083675 Length: 22 February 15, 2002 13:31 Type: N Check: 8143 ..
Found using 'seq4' (pappu403.key)
1 TGACTGTGAACGTCGAGATGA
9 16
-----
1 match found in sequence:
ax083676 ; TOIG of: ax083676 check: 8058 from: 1 to: 22
(from "immun_ge.seq")
TOIG of: ax083676 check: 8058 from: 1 to: 22
LOCUS AX083676 22 bp DNA PAT 28-FEB-2001
DEFINITION Sequence 2 from Patent WO0112223.
ACCESSION AX083676
VERSION AX083676.1 GI:13185408
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 22)
AUTHORS van Nest,G.
TITLE Methods of modulating an immune response using immunostimulatory s
equences and compositions for use therein
JOURNAL Patent: WO 0112223-A 2 22-FEB-2001;
Dynavax Technologies Corporation (US)
FEATURES Location/Qualifiers
source 1..22
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic construct"
BASE COUNT 6 a 4 c 7 g 5 t
ORIGIN
AX083676 Length: 22 February 15, 2002 13:31 Type: N Check: 8058 ..
Found using 'seq4' (pappu403.key)
1 TGACCGTGAACGTCGAGATGA
9 16
-----
1 match found in sequence:
ax083677 ; TOIG of: ax083677 check: 9493 from: 1 to: 23
(from "immun_ge.seq")
TOIG of: ax083677 check: 9493 from: 1 to: 23
LOCUS AX083677 23 bp DNA PAT 28-FEB-2001
DEFINITION Sequence 3 from Patent WO0112223.
ACCESSION AX083677
VERSION AX083677.1 GI:13185409
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 23)
AUTHORS van Nest,G.
```

```
TITLE Methods of modulating an immune response using immunostimulatory s
equences and compositions for use therein
JOURNAL Patent: WO 0112223-A 3 22-FEB-2001;
Dynavax Technologies Corporation (US)
FEATURES Location/Qualifiers
source 1..23
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic construct"
BASE COUNT 6 a 8 c 3 g 6 t
ORIGIN
AX083677 Length: 23 February 15, 2002 13:31 Type: N Check: 9493 ..
Found using 'seq4' (pappu403.key)
1 TCATCTCGAACGTTCCACAGTCA
7 14
-----
2 matches found in sequence:
ax083679 ; TOIG of: ax083679 check: 5564 from: 1 to: 26
(from "immun_ge.seq")
TOIG of: ax083679 check: 5564 from: 1 to: 26
LOCUS AX083679 26 bp DNA PAT 28-FEB-2001
DEFINITION Sequence 5 from Patent WO0112223.
ACCESSION AX083679
VERSION AX083679.1 GI:13185411
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 26)
AUTHORS van Nest,G.
TITLE Methods of modulating an immune response using immunostimulatory s
equences and compositions for use therein
JOURNAL Patent: WO 0112223-A 5 22-FEB-2001;
Dynavax Technologies Corporation (US)
FEATURES Location/Qualifiers
source 1..26
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic construct"
BASE COUNT 5 a 9 c 4 g 8 t
ORIGIN
AX083679 Length: 26 February 15, 2002 13:31 Type: N Check: 5564 ..
Found using 'seq4' (pappu403.key)
1 TCCATACGTCGCGCTAACGTCGTC
6 13 17 24
-----
1 match found in sequence:
ax105150 ; TOIG of: ax105150 check: 5318 from: 1 to: 20
(from "immun_ge.seq")
TOIG of: ax105150 check: 5318 from: 1 to: 20
LOCUS AX105150 20 bp DNA PAT 30-APR-2001
DEFINITION Sequence 48 from Patent WO0122990.
ACCESSION AX105150
VERSION AX105150.1 GI:13921300
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 20)
AUTHORS Hartmann,G.D., Bratzler,R.L. and Krieg,A.U.
TITLE Methods related to immunostimulatory nucleic acid-induced
interferon
JOURNAL Patent: WO 0122990-A 48 05-APR-2001;
```

Coley Pharmaceutical Group, Inc. (US) ; UNIVERSITY OF IOWA RESEARCH
FOUNDATION (US)

FEATURES
source
1. .20
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic Oligonucleotide"
4 a 7 C 3 g 6 t

BASE COUNT
ORIGIN

AX105150 Length: 20 February 15, 2002 13:32 Type: N Check: 5318 ..
Found using 'seq4' (pappu403.key)

1 ATCGACTCTCGACGTTCTC
10 17
|-----|

1 match found in sequence:
ax148636 ; TOIG of: ax148636 check: 8143 from: 1 to: 22
(from "immun_ge.seq")
TOIG of: ax148636 check: 8143 from: 1 to: 22

LOCUS AX148636 22 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 1 from Patent WO0135991.
ACCESSION AX148636
VERSION AX148636.1 GI:14347254
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.

REFERENCE 1 (bases 1 to 22)
AUTHORS Tuck,S. and van Nest,G.
TITLE Immunomodulatory compositions containing an immunostimulatory
sequence linked to antigen and methods of use thereof
JOURNAL Patent: WO 0135991-A 1 25-MAY-2001;
Dynavax Technologies Corporation (US)

FEATURES
source
1. .22
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic construct"
6 a 3 C 7 g 6 t

BASE COUNT
ORIGIN

AX148636 Length: 22 February 15, 2002 13:12 Type: N Check: 8143 ..
Found using 'seq4' (pappu403.key)

1 TGACTGTGACGTTTCGAGATGA
9 16
|-----|

1 match found in sequence:
ax148637 ; TOIG of: ax148637 check: 8058 from: 1 to: 22
(from "immun_ge.seq")
TOIG of: ax148637 check: 8058 from: 1 to: 22

LOCUS AX148637 22 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 2 from Patent WO0135991.
ACCESSION AX148637
VERSION AX148637.1 GI:14347255
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.

REFERENCE 1 (bases 1 to 22)
AUTHORS Tuck,S. and van Nest,G.
TITLE Immunomodulatory compositions containing an immunostimulatory
sequence linked to antigen and methods of use thereof
JOURNAL Patent: WO 0135991-A 2 25-MAY-2001;
Dynavax Technologies Corporation (US)

FEATURES
source
1. .26
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"

source
1. .22
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic construct"
6 a 4 C 7 g 5 t
BASE COUNT
ORIGIN

AX148637 Length: 22 February 15, 2002 13:12 Type: N Check: 8058 ..
Found using 'seq4' (pappu403.key)

1 TGACCGTGACGTTTCGAGATGA
9 16
|-----|

1 match found in sequence:
ax148638 ; TOIG of: ax148638 check: 9493 from: 1 to: 23
(from "immun_ge.seq")
TOIG of: ax148638 check: 9493 from: 1 to: 23

LOCUS AX148638 23 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 3 from Patent WO0135991.
ACCESSION AX148638
VERSION AX148638.1 GI:14347256
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.

REFERENCE 1 (bases 1 to 23)
AUTHORS Tuck,S. and van Nest,G.
TITLE Immunomodulatory compositions containing an immunostimulatory
sequence linked to antigen and methods of use thereof
JOURNAL Patent: WO 0135991-A 3 25-MAY-2001;
Dynavax Technologies Corporation (US)

FEATURES
source
1. .23
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic construct"
6 a 8 C 3 g 6 t

BASE COUNT
ORIGIN

AX148638 Length: 23 February 15, 2002 13:12 Type: N Check: 9493 ..
Found using 'seq4' (pappu403.key)

1 TCATCTCGACGTTCCACAGTCA
7 14
|-----|

2 matches found in sequence:
ax148640 ; TOIG of: ax148640 check: 5564 from: 1 to: 26
(from "immun_ge.seq")
TOIG of: ax148640 check: 5564 from: 1 to: 26

LOCUS AX148640 26 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent WO0135991.
ACCESSION AX148640
VERSION AX148640.1 GI:14347258
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.

REFERENCE 1 (bases 1 to 26)
AUTHORS Tuck,S. and van Nest,G.
TITLE Immunomodulatory compositions containing an immunostimulatory
sequence linked to antigen and methods of use thereof
JOURNAL Patent: WO 0135991-A 5 25-MAY-2001;
Dynavax Technologies Corporation (US)

FEATURES
source
1. .26
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"

gene complement(4472. .4969)
/gene="sfp"
CDS complement(4472. .4969)
/gene="sfp"
/codon_start=1
/transl_table=11
/product="surfactin"
/protein_id="CAB12165.1"
/db_xref="GI:2632658"
/translation="MKIYGIYMDRPLSQEENRPMFTFISPEKREKCRFFYHKEDAHT
LGGVLVRSVISQYQDKDIRFSTQEYKPCIPDLPDAHFNISHSGRWVIGAFDSQ
PIGIDIEKTPISLEIAKRFSTKTEYSDDLAKDKDEQTDYFYHLWSMKESFIKQGRQ
LIASA"
gene complement(5053. .5721)
/gene="yczE"
terminator complement(5053. .5071)
/gene="yczE"
CDS complement(5074. .5721)
/gene="yczE"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12166.1"
/db_xref="GI:2632659"
/db_xref="SPTREMBL:O34927"
/translation="MQELVLRWTFYFAGLIIILAFGVSLTIEGKALGISPDWAFHYSL
FQHGLVQWSIIGALIVGFTSLFRAWPKIGALLNMVLIGVDFDFNFILPAPST
YTSIIVFSGVLLIGVGVVYAGLGAGPRDSLMLLITERKTGNVQVWRNGMEITI
LFAWMGGPIGFGTILTAILTLGLILRLFSLPQSILNLNVAVARRTAAKASPPVH"
gene /gene="yckI"
terminator complement(5815. .5830)
/gene="yckI"
CDS complement(6039. .6782)
/gene="yckI"
/function="unknown"
/note="similar to glutamine ABC transporter (ATP-binding
protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB12167.1"
/db_xref="GI:2632660"
/db_xref="SWISS-PROT:P39456"
/translation="MLTVGLNKSFGENEILKIDMKIEKGVIALILGSGSGKTTLL
RCLNLEIFNRGELAFDDFSIDFSKKVKQADILKLRKSPWVFQAYHLPHRTALENV
MEGVQVQRNKEEVRKEAIQLLDKVKLQDKMDLVPQLSGGQQQQRVGARALAIQPE
LMFDEPTSLADPELVGEVLKVIKDLANEGWTMVVVTHEIKPAQEADEVIFIDGVI
VEQPPPEQIFSAPKEERTQRFNRLNPL"
gene complement(6796. .7500)
/gene="yckJ"
CDS complement(6796. .7500)
/gene="yckJ"
/function="unknown"
/note="similar to glutamine ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB12168.1"
/db_xref="GI:2632661"
/db_xref="SWISS-PROT:P42200"
/translation="MFLNNLPALTILGTAIPMDLVQGSFPIISGGIYYTIPTILTSFI
FCMILALITALARKSVRPLRWVFSVYSAIRGTPLLVQLELIIFYLFPAPNVLDPFP
SAVIAFLNVGAYASEILIRASILSPKQGWAGYTIQHTKTLRVLDPQAFRVSIP
PLSNFTFISLTKLSIASQILVAELFRKAQEGARNLDQILVIYIEAAFIYIICFLLS
LVQVIERLDRYAK"
gene complement(7487. .8293)
/gene="yckK"
CDS complement(7487. .8293)
/gene="yckK"
/function="unknown"
/note="similar to glutamine ABC transporter
(glutamine-binding protein)"
/codon_start=1

/transl_table=11
/protein_id="CAB12169.1"
/db_xref="GI:2632662"
/db_xref="SPTREMBL:P94402"
/translation="MKKALLALFMVSVTAALAACGAGNQNOSKDNQKGDGLWASTKKK
GVLVTGTEGYEPTYHDKDQTKLTGYDVEVITEVAKRLGLKVDKFKETQSGMEAGLN
SKREDVAVNOVGKTRDREKXYDFSCKYITSRAVVVTKKDNNDIKSEADVKGKTSQSLLT
SNYNKLATNAGAKVEGVEGMAQALQIOQARVDMTYNDKLAVLNLKLTSGNKNYKIAF
ETGEPOSTYFTFRKSGSEVVDQVNKALKEMKEDGTLISKKNWGEDVSK"
gene complement(8408. .9280)
/gene="yclA"
CDS complement(8408. .9280)
/gene="yclA"
/function="unknown"
/note="similar to transcriptional regulator (LysR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12170.1"
/db_xref="GI:2632663"
/db_xref="SPTREMBL:P94403"
/translation="MDIRQLRYFITTAQEOKITSAKKLHMAOPPLSRQLKQLEDELG
VVLFRNKKQMTLTYEGAVFLKRAKEILHRFEDAVTEVOELKEEVAGTILAVGSTIYC
AALMLEKVTOIKERYPHLTENIWENEPATLELLESQIDAAVTTTLIKSDTVQFKOL
DDIPCVLIVSDEAGYPCGDTIKWVDIPSPFLLILLRPVNGKGVYNQVWNEHRLNLEPR
IVCECHDSATLLSLVSSGGASILPVTMIPVHMNRNHVHTVIENNPPFIMTPAVMWRD
SYLPKPAQOQFLDLF"
gene 9370. .9984
/gene="yclB"
CDS 9370. .9984
/gene="yclB"
/function="unknown"
/note="similar to phenylacrylic acid decarboxylase"
/codon_start=1
/transl_table=11
/protein_id="CAB12171.1"
/db_xref="GI:2632664"
/db_xref="SPTREMBL:P94404"
/translation="MKAFFKRGGKVKLVGMGTGATGATFGVRLQLQWKAAGVETHL
VYSPNANVTIKHETGYTLQVEQLATYTSHKDQAAAISSGSPDTGDMIVAPGSMKSL
ASIRTGADNLTLTRAADVMLKERKLVLLTRPTPLQIHLNMLALTCKMGTIILPPMP
AFYNRPRSLEEMVDHIVFTLDQFGIRLPEAKRWNGIEKQKGA"
gene 9987. .11408
/gene="yclC"
CDS 9987. .11408
/gene="yclC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12172.1"
/db_xref="GI:2632665"
/db_xref="SPTREMBL:P94405"
/translation="MAYQDFREFLAALEKEGOLLTVNEEVKPEPDLGASARAASNLGD
KSPALLFNNIYGHNARIAMNVIGSWPNHAMLMKPKDTPVKEQDFEFKRYDQFPMP
VREETAPHENEITDINFLDILPRLNQDGGYILDKACVISKDLDDPNFGKON
VGIYRMQVKGDRLGIQVPQHDIAIHLRQAEERGINLPTLTALGCEPVITTAASPTL
LYDQSEYENAGAIQGEYRIVKSKLSLDLPWGAEVYLEGIIAGREYEGPGEFTG
HYSGGRSMPIIKIRVYHRRNPFIHLYLGMPTWECYMGINTCVPLVQOLKEAYPN
EIVAVNMYTHGLIATVSTKTRYGGFAKAVGRALTTPHGLGYCKMWIVDEDVDPFN
LPQVMALSTKMKPHKDAVIIPDLVLPDPSNSGITHKMLILDATTTPVAPETRGHY
SQPLSPSLTTKEWEKMLDLMNK"
gene 11425. .12140
/gene="yclD"
CDS 11425. .12114
/gene="yclD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12173.1"
/db_xref="GI:2632666"
/db_xref="SPTREMBL:P94406"
/translation="MHTCPDCSKKGEVMSKSPVEGAMEVYOCOTCFFTWRSCEPESI
TNPEKYNPAFKIDPKETETAIEVPAPPEPKGLIRVNCMSDRLELLDGLSCLNEKQHEA


```

/codon_start=1
/transl_table=11
/protein_id="CAB12181.1"
/db_xref="GI:2632674"
/db_xref="SPTREMBL:P94411"
/translation="MSLQFOQVGYWYKNSQPLFODINISFOKGFYTIIVGTSGTGK
TTFLSLAGLDAPKEGNIYDGKAVSKIGITNFRNOYVSIVFQAYNLLPYMTALQWYT
TAMEITGSEKKNKESVALDMLQKVGNEKQARQKVLTLSGGQOQVSIITRAFCCDDTL
IVADEPTGNLDEDTSEIIVLFDLAHKEDKCVIMVTHDEQIAKVSIDINIRLSRGSFT
VKENAVV"
complement(21734..23194)
/gene="yclI"
/gene="yclI"
/complement(21734..23194)
/function="unknown"
/notes="similar to transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB12182.1"
/db_xref="GI:2632675"
/db_xref="SPTREMBL:P94412"
/translation="MNFIKRAFNMKAKGKTKLLQLFVFTVICFVLSGLAIOASQAQK
SSELAQELGGSVTIQVDQKQMEKQDSGERKTFESTPIKVSIDANKLAALDHWKSYN
YTTASANAGNDAIESSSSDSSSSNAKNSQGGGQGGQPMQVADLSTEGVIST
ALVDDFSQDSKITDGRAITKSDGKVTVINETLAENDLSVGDSITIESATDEDTT
VKLIVGIVKTTSSGDDQANFSFLPNKLYPTATPAALGGDDKNTIDSAYITMD
DAKNMDFVKAAKKTSIDDTITLNTDOLYQOMVGPENVASFSKNVYLVSVAGAV
IGLIVMSIRKRYEMGVLMAGIEKRWKLIQGFLEILIVAVIAGLASVTGNLVAN
QIGNQLLSQISSDSTQTASQMPGGGGGGMGRMFHSSSNVDVIDSLNVAVSMDN
MLILGIGILIAIATLPLSPISVLRLHPKTLTKQE"
23407..24090
/gene="yclJ"
/gene="yclJ"
/23407..24090
/function="unknown"
/notes="similar to two-component response regulator [YclK]"
/codon_start=1
/transl_table=11
/protein_id="CAB12183.1"
/db_xref="GI:2632676"
/db_xref="SPTREMBL:P94413"
/translation="MKILMIEDNVSVCTMTMEFFFEKGFGEAFVHDGLGEGYORFTEEN
WDLIILDMPSMDGVTICRKIRETSTVPIIMLTAKDTESDQVIGFEMGADYVTKPF
SPLTLVARIKAVIRRYKATGKAVIDEDMLECFITINKTREVLLNGEPVENLTKEF
DLIYLVONPQOVFSRQLLEQVWGVIQYIGDRTVDVHIKRLKRLKLAESDKPFLYTVW
GVGKAFDSB"
Join(24077..25498,25506..25531)
/gene="yclK"
/24077..25498
/gene="yclK"
/function="unknown"
/notes="similar to two-component sensor histidine kinase
[yclJ]"
/codon_start=1
/transl_table=11
/protein_id="CAB12184.1"
/db_xref="GI:2632677"
/db_xref="SPTREMBL:P94414"
/translation="MMKIKLYQLLSHISILILAFVIIISLFSHFVREFAYQNKVEE
LTSXAVQIANEFQSQVDMRLYPYQDILSTRKTOFIIFNEEQPYFLPEGFHPHPE
QLKSEWNKLGKQVTVTRADCRFDEVSVAOPFVQNEFKGAVLLISPLSGVEQMV
NOVNLNMFYAVISTVLITILVSWLLSKPHVKRIQKLREATDKVASGDDTDIHLNSYGD
EIGVLASDFNMAKLLQSRDEIRLEKRRQFTADVSHELKTPITTINGLEVLNSH
TIPQDKDCFSLSIEQDTRMLRVKENLDYERKSQOITLTKLDVPLIEVFEIVKEH
LOQAEERQKMLQIODEHVIVHADYDFRFTQILVNTKNSIQFTQNGDIWLRGMGEYK
ETIIEIDTIGICIPKEDIEHIWERFYKADISRTNTAYGEYGLGLSIVRQLVEMHQGTV
EIKSEEGKTKFIIRLPLTAQK"
25506..25531
/gene="yclK"
/25661..26809
/gene="rapC"
/25661..26809
/gene="rapC"

```

terminator

gene

CDS

```

/notes="alternate gene name: yclL"
/codon_start=1
/transl_table=11
/product="response regulator aspartate phosphatase"
/protein_id="CAB12185.1"
/db_xref="GI:2632678"
/db_xref="SPTREMBL:P94415"
/translation="MKSQVIPSASVGQKINEWRYIRTFSPDAEVLKAEIQOELKHM
QHDNLLYYSLMEFRHOLMDYVLEKLNIEDQPSLSLSRNDISNOADLGLLDY
YVNFEGMYEEDKREFTSAITYKQAEKLSFVADHIERAEFYFKIAEAYYMKQTFYF
SLINIKNAYEIIYVEQETYNRIICHPFVGVNLMERNFEQAAHFALNMAQEQK
AQDGRAYYNLGLCYNQDULDDPAIDYFKAVSFTFESSRIVNSLPQAYFLITLIYKO
GRHKDASEYHKRGYEAFTDDADYAVKFFLQSLYLDQPNEEGIERCFQYLNKNMY
ADIEDLAEVAKYIYEQKWEKLSASVFLQVFEARKQQRSEGLYEIEI"
Join(26793..26915,26923..26946)
/gene="phrC"
/26793..26915
/gene="phrC"
/function="CSF stimulates competence gene expression
(srfA) at low concentrations, inhibits competence gene
expression at high concentrations, and stimulates the
ability of cells at low cell density to sporulate at high
concentrations"
/codon_start=1
/transl_table=11
/product="regulator of the activity of phosphatase RapC
and competence and sporulation stimulating factor (CSF)"
/protein_id="CAB12186.1"
/db_xref="GI:2632679"
/db_xref="SPTREMBL:P94416"
/translation="MKLKSFLVICLAAAIFTAAGVSANAEALDFHVTGRGMT"
26923..26946
/gene="phrC"
/complement(27421..28817)
/gene="yclM"
/complement(27421..27440)
/gene="yclM"
/complement(27453..28817)
/gene="yclM"
/function="unknown"
/notes="similar to homoserine dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB12187.1"
/db_xref="GI:2632680"
/db_xref="SPTREMBL:P94417"
/translation="MKVVKFGSSLAGQLDKVFHIVTSDPARKAVVVSAPCKHYAE
DTKVTDLIIACAEQVLTGSAPELAAYVERVALIANELQSGSIIETKIRDDLTLE
GKSNPEQYLDVAKSGEDNNAKLIATAYFRYKGVKAEYVNPDKAGLVTFNPGNAQVL
PESYONLYRLRERDGLTIIPFPFGFGSKDGVITFSRSGSDITGSLANGLOADLYENF
TVDVAVSYNPSFVENPKEISELTREMRELSYAGFSVFDEALIPAFRAGIPVQIKN
TNPSAEGTRVYKRDNTNGTVPVVGIASTDGFCGSIYISKYLMNRICGFRRALQILEEH
GLTYEHVPSGIDDMTITLRQOMDAATERSVIKRIEFDLHADEVIVVEHIALIMVVEG
AMRHNVGITARAAKALSEAQVNIEMINQSSSEVSMFGVKEAEERKAVQALLQEFFAG
VLIS"
29202..30152
/gene="yclN"
/29202..30152
/gene="yclN"
/function="unknown"
/notes="similar to ferrichrome ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB12188.1"
/db_xref="GI:2632681"
/db_xref="SPTREMBL:P94418"
/translation="MKRLYFLILLIILAVTSVFIGVEDLSPLDLSKQEAFTLSPAS
RLPRLISIVLAGLSICGLIMQISRNKFSPTTAGTMDWARLGLTLLISLFTSASP
LTKMLFAVFAAGLVNFKILERIKFNDFIPLVGLMGNIVSIAFTIAYKYDLI
QNSWMLQDGSFLVGRVLYLSPILVIIVAVVADKFTLAGMGESFSVNLGLKYKR
VYNIGLIIVTSILVILTVGLMPLGLIIPNIVSYIRGDNLKSSLPHVTVLLGAVFVL
FCDILGRILIEPYEISIGLWGLIGSGIFLEMLLRKAYA"
30145..31092

```

gene

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------|--------------|---------------|---------------------|---|----------------|------------------|--------------------------|-----------------------|----------------------------|---|---------------|--------------|---------------|--------------|---------------------|--|----------------|------------------|--------------------------|-----------------------|----------------------------|---|---------------|--------------|---------------|--------------|---------------------|--|----------------|------------------|--------------------------|-----------------------|----------------------------|---|---------------|--------------|---------------------------|--------------|---------------------------|--------------|---------------------|---|----------------|------------------|--------------------------|-----------------------|----------------------------|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| CDS | /gene="yc10" | 30145. .31092 | /function="unknown" | /note="similar to ferrichrome ABC transporter (permease)" | /codon_start=1 | /transl_table=11 | /protein_id="CAB12189.1" | /db_xref="GI:2632682" | /db_xref="SPTREMBL:P94419" | translation="MRNQMKIALLVGLAIVICIGLFLFYDLGNWDYTLPRRIKKVAAIV LTGATAFSMIQOTINNRIILPSILGLSLYMLIQTGIFLPGSNWIMNNKINIF IISVLLMLFSLVLYQIMFGEGRNIFILLIGIVITFLFSSLSSEMQMLIDPNEFQV VQKMFASFNNINDLLWLAFIIFLPLRGVYVWRFTKFDVLSLGRHNAVNLGIDYKV VKQMLIVVAILVSVSTALVGPIMFLGLLVNLAAREFLKTVKHSYLVAFISIIALV GGQFVVEKVTFTSTLSVIINFAGGIYFIVLLKENKSW" | 31086. .31844 | /gene="yc1p" | 31086. .31844 | /gene="yc1p" | /function="unknown" | /note="similar to ferrichrome ABC transporter (ATP-binding protein)" | /codon_start=1 | /transl_table=11 | /protein_id="CAB12190.1" | /db_xref="GI:2632683" | /db_xref="SPTREMBL:P94420" | translation="MVEVRNVQYGGKVVLEETSVTIQKGKITSFTGPNAGAGKSTLL SIMSLRIKDSGEIYIDQEIAGCDSKELAKKWSILQANQINIRLTIKDLVSFGRFP YSGNRLTEEDWHINQALSVMKLEIDQKYIDQLSGGCQCORAFIAMYIAQDTYIFLD EPLNLDKHSVSEIMKLLKLRLVEELGKTIIVIVIHINFASVSYDYIVALKNRIVKGG PPEMIETSVLEEIYDWTPIQTIDNRIGVFS" | 31866. .32854 | /gene="yc10" | 31866. .32819 | /gene="yc1q" | /function="unknown" | /note="similar to ferrichrome ABC transporter (binding protein)" | /codon_start=1 | /transl_table=11 | /protein_id="CAB12191.1" | /db_xref="GI:2632684" | /db_xref="SPTREMBL:P94421" | translation="MKKFAELLFTALVTAVVISACGNGNSTSSKSGDTKKKEQITVKHOLD KNGTKVPKNPKVWVDFGSLDTLDKGLDDIVAGLPKQVLPKLYLSFKDDKYADVGS LKPEPDFKVAELDPDLIIISARQSESYKEFSKIAPTIYLGVDYAKYMESFKSDAETIG KIFDKEDKDELANDIHSIADYKTKAEKLNKGLVIMANDGKISAFGPKSRYGLIHD VFGVAPADQNIKASTHGQSVSYEYISKTNPDYLFVIDRGTAIGETSSTTKQVVDYVK NVNAVKNHGVLYLDSATWYLSGGGLESMTQMIKEVKDGLEK" | 32829. .32854 | /gene="yc1q" | complement(32866. .34284) | /gene="ycnB" | complement(32866. .34284) | /gene="ycnB" | /function="unknown" | /note="similar to multidrug resistance protein" | /codon_start=1 | /transl_table=11 | /protein_id="CAB12192.1" | /db_xref="GI:2632685" | /db_xref="SPTREMBL:P94422" | translation="MNTSIEQKPFNRSVIVGILLAGAFVAILNOTLTLITLPHIMRDRF NVDAQOQWLTTSFMLTNGILIPITALEIKFTSRALLITAMSIFFTAGTVVGFAPNF PVLITARIIOAAGAIMPLMQVFLTFIPTEKRGQAMGVGLVISFAPAIQPTLSGW AVAFSWSRSLFYIILPPAVIDLILASILMKNTVTLRKTDIILSVILSTFGFGLLYG FSSVSGYSGWSSTVLISLLVGVIALLFIITROMKLLKPMLEFRVFTGVSFLTLLCT LVFALLIGITETILPLYTONVRDVTAFDTGLMLPGAVVMGFMSPIIGRIIDRVGGRGL AIFAGCLIFLTSPLFMOLTDBTSLAWIVLIVTYVRLGTAMIMMPVTITAGINALPRHLI PHGTAMNNTIRQVGSGTALLVSMNSQAAHAGTTNVKHAALHGMNAAFIVAAVIAL VGFLLSFTLKKPQPAEQPAR" | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| gene | terminator | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

/db_xref="GI:2632690"
/db_xref="SPTREMBL:P94426"
/translation="MDITITLDRSEQADYIYQIYQKLLKRNLLPHSKVPSKRE
LAELKYSVNSGAYQOLLAEGYIAIERQKAEVSEELDMFSAEHPFALPDDLKEI
HIDQSDWISFSSMSDDHPPIKSWFRCEKAAASRSTRTGLDMSHPGQIYEVRAITR
LISLTRGVKCRPEOMIIGACQTVLMQLLTLLPKEAYAMEEPGYRRMYQLLNAGKQ
KTTMLDEKMSIAETIROQPDVLVITTPSHQFPSTIMPVSRRIOLNWAAEPPRYI
IEDDYDEFTYDVDSIPALQSLDFQNVYMGTFESKLLPGLRISYVMVPELLRAYK
QRYDLQTSLSLTQLTQEFIESGEYQKHKKMQHYKEKRELRITALEAFSGEVTV
KGAGLHFVTFEFTRRTEQDILSHAAGLOLEIFGMSRNLKENKROTGRPALLIGFA
RLKEEDIQGVQRLFKAVYGHKKIPVTGD"
complement(36859. .36874)
/gene="yncF"
complement(36859. .36874)
/gene="yncF"
38401. .39711
/gene="yncG"
38401. .39711
/gene="yncG"
/function="unknown"
/note="similar to 4-aminobutyrate aminotransferase"
/codon_start=1
/transl_table=11
/protein_id="CAB12198.1"
/db_xref="GI:2632691"
/db_xref="SWISS-PROT:P94427"
/translation="MSQTTASITTAQWQKRDQFVSKGVSNGNRSIAVKGEGAEYDL
DGRSIFDEAGIAGTINVGHSKPKVAVKRAQAEELIHPGFNMVMPYIIEAEKLCGI
APGSHKKAIFLNSGAEAVNAVKIARKYTKROGVSTFRGFHGRNTMTKSMTSKVKP
YKFGFGPAFVQAPFPYVYQKPGAGMSDEYDDMVIOAENDFIASVAPETVACVM
EPQGGEGFIPSKRFVQHVASCKENGIVFADEIQTGFARTGYTFAIEHFDVDPOL
ITVSKSAGALPLSGVIGRAEMDAAAPGELGGYAGSLGCAAAALAVLDIIEEGLN
ERSEIKGILIEDKAYEWKQEPFFGIDIRLIGMAAAIEIVKDPDTPREDPKTKAAIAAY
ANQNGLLLAGINGNTIRFLTPLVIVSDLSLNEGLSILEAGLRA"
39780. .41211
/gene="yncH"
39780. .41168
/gene="yncH"
/function="unknown"
/note="similar to succinate-semialdehyde dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB12199.1"
/db_xref="GI:2632692"
/db_xref="SPTREMBL:P94428"
/translation="MPDLTVYPATGEEIKTIPOQSAATEVEEAIERSHOAFKTSKT
SANERTSLKWEYELIVEHKEELADLITKENGKPYQEAUGEVLVYAGYIEWFAEEAKR
VYGRTPAPTGTGRIVVTROQVGPVAAITPWNPNAMITTKAAPALAAAGCTFIKPPAP
DTPLSAYELARLAVEGIPKDVQVIGDGEIGCNVTSPPKIRKIFTGSTVPVKIL
MKNSADTVKIVSMELGSHAPLIVDEADIDILAVEQAAASKYRNAGQTCVCANRLVHE
SIKDEFAAKLUSEQVSKLVNGLEGVNVGPIINRGFEKIVISOIDDAVEKGAKVITAG
GTYDRNDKGCYFVNPTLVTDVDTSMNIMHEETFGPVAPIVTFSDIDEATQLANDTPY
GLAAYFTENYRRGIYISENLEYIIGWNDGGSAPVQAPFGGKESGIGREGGSEGIE
PYLETKVLSTGL"
41187. .41211
/gene="yncH"
41291. .42154
/gene="ycxE"
41291. .42154
/gene="ycxE"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12200.1"
/db_xref="GI:2632693"
/db_xref="SPTREMBL:P94429"
/translation="MDLLALLPALFWGSIVLVNVLGGGPPYSQTIGTTGALIVSIV
IYFFVPLSLRIFIVIGVSLGWSGOANQLKSIOLMGVSKTNPISGTQMLVSTILF
GVTFREWSTVLAIITGLVLAIFIIVGIIILTSLEKNDKKEGSPNLKKGILLVYST
LGVLVVVARLENVSGWSALLPQALGMVGVGLVLYRHKFPNKYAIRNLPLGLIWAG
GNMELFISOPRGVATSFSLSQMGIVISTLGGFIFILREKTKRQLIAIGIILIAA
AVFLGIATKNS"

gene
CDS
42174. .42959
/gene="gdh"
42174. .42959
/gene="gdh"
/EC_number="1.1.1.47"
/codon_start=1
/transl_table=11
/product="glucose 1-dehydrogenase"
/protein_id="CAB12201.1"
/db_xref="GI:2632694"
/db_xref="SPTREMBL:P94430"
/translation="MYPDLKGVVAITGAASGLGAKAMAIRFGKEQAKVIVINYSNKQD
PNEYKEEVIRKAGBAVVQGDVTKKEEDYKNIVQTAIKEFGTLDIMINNAGLENVPVPSH
EMPLKMDKVICTNLTGAFGLSREAIKYFVENDIKGNINNSKVEHVIPWPLFVHYAA
SKGIGKLMTEITALEYAPKGRVNNIGPAINTPINAEKFPADPKQKADVESMIPMGYI
GEPEIAAVALASKEASYVTGITLFDAGGMTQYPSFQAGRG"
complement(42969. .43618)
/gene="ycnI"
complement(42969. .42995)
/gene="ycnI"
complement(43004. .43618)
/gene="ycnI"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12202.1"
/db_xref="GI:2632695"
/db_xref="SPTREMBL:P94431"
/translation="MLKKIALTLPATVGSLLFTAPASAHVSVKPAESAAGSHETYYT
MKVSEKNLPTTKVLKMPKDEVEQQEYPIPGWKVSTQKHDDKSVSVTWEATDGGIQE
GQOQFTFVAKNPKABEAAWDAYYKDGSIIVEMTGDDEADTPHSITNITTSKQVTD
EHGATKTEDSENKSGSALDITAMVLSAAAILSVAAALVKKKRA"
complement(43631. .45256)
/gene="ycnJ"
complement(43631. .45256)
/gene="ycnJ"
/function="unknown"
/note="similar to copper export protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12203.1"
/db_xref="GI:2632696"
/db_xref="SPTREMBL:P94432"
/translation="MKRNRMWIILLFLVFLPKTSFAHAYIVKSSPGENSELKSAQAQ
VEYENFPVBEGFHYIKVNSNGRDLDTTEIKKDNHHIMTVLKNLKPDKDVRAEW
NAVSDGHPVSGVYFPFSGRADGFSOKAADSALNPGTAADRAILYLTALSLFTGTVF
FHLFWYKSEOLVKRTRRILITGSIAALGLALLQLPIOTKANAGGWSAFQPCYIR
ETLFEETAGGSIIQAAFLVLLALSIVIPAIRKNRFSFGYWTAPLIFPFGLLAKAFT
GHAAVVEKVTVGILMDLHUTSASIIWVGGAALVLLLSKEMWQPKTLAWETVRRFSP
WALTAVGILVPSGLLNGFFIIRSDMSLFTAYGQALLVYKGLFVMLVGLGAIHFLLR
KQRTTGISRTLKAEMWIGIAVLITAAVFTSLPSPPEPAPEFYQTKAENGQSVLSI
SPNQPKNVFELRVTDHNGDPVKNIQOITLTVYKTGLSGSENKSTFTTLKTKTGVFQD
ONLSINEKGNWKIKVHGLTGDENEINIMFTKN"
complement(45291. .45863)
/gene="ycnK"
complement(45291. .45863)
/gene="ycnK"
/function="unknown"
/note="similar to transcriptional regulator (DeoR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12204.1"
/db_xref="GI:2632697"
/db_xref="SPTREMBL:P94433"
/translation="MLPINRQOHLKLWKBEGSLRISDISARFCGVSEMTVYRDVNQLV
QSNQIKTAGTILPVRTPQTDHMCSTYCLKPVNQAHSVQLITVQNQDIEQLCCAHCAFU
RYADKTEEVSHLICRDFLQTTVSAGSAYFVNAELNHCCQPOAIPFATILDHAERFQ
KFGCGAVCTFDQALDMLQDRKKRCTCKR"
46028. .46392
/gene="ycnL"
46028. .46381
/gene="ycnL"
/function="unknown"

terminator
/codon_start=1
/transl_table=11
/protein_id="CAB12205.1"
/db_xref="GI:2632698"
/db_xref="SPTREMBL:P94434"
/translation="MKETPCPCGKPLTGDVMYRSSNVPCQFCGCHCRERLYEYKVSAP
IMVLSAAIVLLIYLLMLRNAAGSVLPVQVHPMAVFAVCAVPFIVSERMIARY
IQGNIIYGRKRKGS"
46373..46392
/gene="ycnL"
4654..48386
/gene="mtlA"
4654..48386
/gene="mtlA"
/EC_number="2.7.1.69"
/codon_start=1
/transl_table=11
/product="phosphotransferase system (PTS)
mannitol-specific enzyme IABC component"
/protein_id="CAB12206.1"
/db_xref="GI:2632699"
/db_xref="SPTREMBL:P94435"
/translation="MQQEQOOGGMKVQVREGSVLSCMIMNIGAFIANGIITALFI
PAGFPMENLTVSPMITLPLLIAYTGGKMIYDHRGCVGATAAIGVIGSDIPM
FIGAMIPGLGYLIKQTDKFKVKQGFGLINNTAGIVGAALTILAFYAIGPV
LTLNKLAAVEVIVHANLLPVASVLEPAKVLFLNNAIHGILSPIGIEASQTGKS
ILFLVANGPGIGIFLAYMFGKSGSSATPGAAIHFPGYIHFYFYLKMQPGD
SPSHCRRSKRTLNITFNAGLVAASPGSIIALMAMTPSGYEGVLAVLVAAYSFI
VSAYILKSAEEDLAATEKQMSKGGKKSQAAALEAQAKERKSLCLKARTKL
SPRVNIDGKCHGGIHLKKSEKAELDITVNTAINLPDADIVITHKDLDRAKK
LGNATHISYDNFLNSPKYDELKLSNLIERYCHASTRGNIKLNQTVSSKEEAIK
LAGQTLIDNGYVEDYISKMFDEETSSTFMGNFIAIPHGTBEAKSEVLHSGSIIQI
PEGVEGENTAKVVFGIAGKNHELDILNSIAIICSEETLNA"
48439..49584
/gene="mtcD"
48439..49539
/gene="mtcD"
/EC_number="1.1.1.17"
/note="alternate gene name: mtlB"
/codon_start=1
/transl_table=11
/product="mannitol-1-phosphate dehydrogenase"
/protein_id="CAB12207.1"
/db_xref="GI:2632700"
/db_xref="SWISS-PROT:P42957"
/translation="MIALHFGAGNIGRGFIGALLHHSGYDVVVFADVNETMVSLLNEKK
EYTVLAEGRSRSEIIGPVSAINSQSOTELRYRLMNEAALIITAVGPNVKLITAPSA
EGLRRNTANTLNIACENMIGGSSFLKKEIYSHLTEAQKSVSEPLGPPNSAVDRIV
PIQHEDPLKVSVEPFPEWIDESGFKGTPIVINGALFYDDLTPYTERKLFTVNTGHA
VTAYVGYQGLKTVKEAIDHPEIRRVVHSALLETGDLVKSYGFKQTEHEQYIKNORS
LLKSPHFGRCDPRSEVTSQTKGKCRLVGPAKKIKEPNALAEGLAAALRFDFTGDPEA
VELQALIEPKDPAAYFKRCAAFSPMNRCTPSP"
49564..49584
/gene="mtcD"
49647..50552
/gene="ycsA"
49647..50552
/gene="ycsA"
/function="unknown"
/note="similar to 3-isopropylmalate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB12208.1"
/db_xref="GI:2632701"
/db_xref="SWISS-PROT:P42958"
/translation="MTMKQFEIAAIPGDGVGKEVVAABEKVLHTAAEVHGGLSFSFTA
FPWSCDYYLEHGKMPEDGIHTLTOFEAYFLGAVGNPKLVPDHIISLWGLLLKIRRELE
LSINMRPAQMAGITPPLLHPNDFVWIRRESEGEYSEVGGRIHRGDDIEAIONAVF
TKKATERVNREAFELAKRRRTLVTATKNSGLYHAMPFWDEYFOOTAADYSGETSSQ
HTDALAAFTVTRPEFDVIVASKLFGDILTDISSLSMESIGIATPPDINPSGKYPMFE
PVHGSAPDIAGQALPIRSARFGQRS"
50842..51510
/gene="sipU"

CDS
50842..51405
/gene="sipU"
/EC_number="3.4.21.89"
/note="alternate gene name: ycsB"
/codon_start=1
/transl_table=11
/product="type I signal peptidase"
/protein_id="CAB12209.1"
/db_xref="GI:2632702"
/db_xref="SWISS-PROT:P42959"
/translation="MNAKTITLKKRKIKITIVVLSIIMIAALIFTIRLVEYKPELTIG
SMAPTLDKDSERILVDKAVKWTGFGHGDIIIVHDKKSGRSFKVRLILGLPGDSIKMKKN
DQLYINDKVEEYPLKEYKQEVKESGVTLTGDFEVEVPSPGIVMFGDNRLNSLDSRNG
MGMPSEDDIIGTESLVFPFGEMROAK"
51485..51510
/gene="sipU"
complement(51515..52072)
/gene="yczH"
complement(51515..52072)
/gene="yczH"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12210.1"
/db_xref="GI:2632703"
/db_xref="SPTREMBL:O31482"
/translation="MRPKPLVILAHETIYGVNSHMKMGRLIKMAGYDVLTPNLLGEDE
VTLKREKTAQEOTKHERLKTGETIIQPIRONAGRHIEVIGFSGCATIAWKCSSMP
EVSGSVYGSRIKSLHHPACVLLFFPNYEPFSDVALLIKLREKQHTHLEIYQF
DALHASPILILFISTGTFPSSKRFLS"
52158..52670
/gene="ycsD"
52158..52670
/gene="ycsD"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12211.1"
/db_xref="GI:2632704"
/db_xref="SWISS-PROT:P42961"
/translation="MNSLSLPHRYPELFDIGVTDSEPGKHAAYKLISENDFWITDQ
TEMPFSLVIEGLRKLRLQELQMTAMDYCLRRKSKWEKLCCLVTGLILRLKLSRAIGAG
LCSDMPRHQLESLLKLSASILKNEKNPSQHVRGVSILCRNPVPRESYLFAGVCVF
HHSYFPPSALH"
52873..53649
/gene="ycsE"
52873..53622
/gene="ycsE"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12212.1"
/db_xref="GI:2632705"
/db_xref="SWISS-PROT:P42962"
/translation="MSVQREDVDIKLIAIDMGFTLLNDEOLISDENKKAITREAEKGV
YVVISGRTLMTCRELAESLKLSSFLITANGSEINSDNFNLVBRKLLHTDHIOMMDEL
RNKHNINFASTVKNVRGEFFENITDHEWLKFGFDIEDDDDIRNEVLEELRKYKSSSI
TNSSTPIEYNALGINKAAAPCGYKGLGFTMENVMAMGSLNDIAMIKREAGLVGANG
NAQDIVKETADYITDNTIEDGVAKAIRHWVL"
53628..53649
/gene="ycsE"
53827..54462
/gene="ycsF"
53827..54462
/gene="ycsF"
/function="unknown"
/note="similar to lactam utilization protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12213.1"
/db_xref="GI:2632706"

CDS
50842..51405
/gene="sipU"
/EC_number="3.4.21.89"
/note="alternate gene name: ycsB"
/codon_start=1
/transl_table=11
/product="type I signal peptidase"
/protein_id="CAB12209.1"
/db_xref="GI:2632702"
/db_xref="SWISS-PROT:P42959"
/translation="MNAKTITLKKRKIKITIVVLSIIMIAALIFTIRLVEYKPELTIG
SMAPTLDKDSERILVDKAVKWTGFGHGDIIIVHDKKSGRSFKVRLILGLPGDSIKMKKN
DQLYINDKVEEYPLKEYKQEVKESGVTLTGDFEVEVPSPGIVMFGDNRLNSLDSRNG
MGMPSEDDIIGTESLVFPFGEMROAK"
51485..51510
/gene="sipU"
complement(51515..52072)
/gene="yczH"
complement(51515..52072)
/gene="yczH"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12210.1"
/db_xref="GI:2632703"
/db_xref="SPTREMBL:O31482"
/translation="MRPKPLVILAHETIYGVNSHMKMGRLIKMAGYDVLTPNLLGEDE
VTLKREKTAQEOTKHERLKTGETIIQPIRONAGRHIEVIGFSGCATIAWKCSSMP
EVSGSVYGSRIKSLHHPACVLLFFPNYEPFSDVALLIKLREKQHTHLEIYQF
DALHASPILILFISTGTFPSSKRFLS"
52158..52670
/gene="ycsD"
52158..52670
/gene="ycsD"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12211.1"
/db_xref="GI:2632704"
/db_xref="SWISS-PROT:P42961"
/translation="MNSLSLPHRYPELFDIGVTDSEPGKHAAYKLISENDFWITDQ
TEMPFSLVIEGLRKLRLQELQMTAMDYCLRRKSKWEKLCCLVTGLILRLKLSRAIGAG
LCSDMPRHQLESLLKLSASILKNEKNPSQHVRGVSILCRNPVPRESYLFAGVCVF
HHSYFPPSALH"
52873..53649
/gene="ycsE"
52873..53622
/gene="ycsE"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12212.1"
/db_xref="GI:2632705"
/db_xref="SWISS-PROT:P42962"
/translation="MSVQREDVDIKLIAIDMGFTLLNDEOLISDENKKAITREAEKGV
YVVISGRTLMTCRELAESLKLSSFLITANGSEINSDNFNLVBRKLLHTDHIOMMDEL
RNKHNINFASTVKNVRGEFFENITDHEWLKFGFDIEDDDDIRNEVLEELRKYKSSSI
TNSSTPIEYNALGINKAAAPCGYKGLGFTMENVMAMGSLNDIAMIKREAGLVGANG
NAQDIVKETADYITDNTIEDGVAKAIRHWVL"
53628..53649
/gene="ycsE"
53827..54462
/gene="ycsF"
53827..54462
/gene="ycsF"
/function="unknown"
/note="similar to lactam utilization protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12213.1"
/db_xref="GI:2632706"

/db_xref="SWISS-PROT:P42963"
/translation="MFQIDLNCDLGSEFGAYKIGLDQDILEYVTSANIACGFGHAGDPS
VMKTVALLAERKVRKGAHPGLPDLGFRGNMAISPPEEAYDLVVYQIAGLSFLKAE
GLHMQVHKPHGLAYNMAVDOKLSDAIAKAVYKVDEGLILFLGLAESELVKAGERIGLQ
TANVEFADITYQSDGTLTPAPSRTHSLKVMMPRHGQNGERRGGQISARP"
54666..55826
/gene="ycsg"
CDS 54666..55826
/gene="ycsg"
/function="unknown"
/note="alternate gene name: ycsH
similar to branched chain amino acids transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB12214.1"
/db_xref="GI:2632707"
/db_xref="SPTREMBL:P94436"
/translation="MGAFLMATSAIGPGFLQTQTATFTNTLAASFGFVILISILDF
AOTNVWRIIAVSGKGOEIANMVLPGLYSIALIVLGLAFNIGNIGGAGLGLOVLF
GITPETGLISAVIALILFVKEAGIAMDRFTQIAGFVMIILTYVVAANTAPPVQAV
ANTFPEHSIFAIVTLVGGTVGGYITFAGGHRLLDAGIKGESIPQVTKSSVVGILI
TSVMRIALFAVLGVVSGKLHIDESNPAASVFKLAAAGNVGYKIFGLIMSAATISVIG
AAYTSVTFEFTSPKIERNSGIIITGFIIVVSTLAFVTIGQPAKILVLVGLSLGLILPI
ALGTLVAAAYKKNIYGVGYKHLPLWLTSTGALVIVMAVMGIYTLCTQLPLWS"
55850..56623
/gene="ycsi"
CDS 55850..56623
/gene="ycsi"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12215.1"
/db_xref="GI:2632708"
/db_xref="SWISS-PROT:P42966"
/translation="MAPKDVRLALIREGKNGPTAGMSGGYAOANLVVLKDLAFDPLL
FCQRNKPCLVDTEANGSPVPSLLRMLLSRTDFPKYRIYRHGILTEYSDITPYWE
DFVFGFLGCSFSGQALINNGIAVHIDEGTNVSMYKTNIDCFPVGAFHGQVMSMR
PYPERLAVRAAOTSRFPVAVHGGPIHIGNPGAIGIDLKPGDFGDVAVSRDGEVPVFW
ACGVTPQAVAMVKNPVMVITHAPGHMLITDIDESLAVL"
56667..58403
/gene="ycsj"
CDS 56667..58403
/gene="ycsj"
/function="unknown"
/note="similar to allophanate hydrolase"
/codon_start=1
/transl_table=11
/protein_id="CAB12216.1"
/db_xref="GI:2632709"
/db_xref="SWISS-PROT:P42967"
/translation="MTVRYQIQGLGDSAMMIRFGEEINEQVNGIVHAAAAYIEEQFPF
GFIECIAPFTSLTVFYDMYEVYKHLPGQISSPFESVKRDEVERLAEIADYEVNRRIV
ETPCYCGEFGDLEVAKINOLSPVEVIDIHTNGEYVVMGLFAPGFPFLGMSKRI
AAPRSSRPSPISAGVIAGLTQVYPISTPGQQLIGKTPLALFRPOENRPTLLRA
GDIVAFRISEKITPIRRSPHEHISIKARTAHNGSGYQNGLPKIRKSGRRYHGHE
TAKOSVDRKRNEAGLEITMMGPSPFHSKOTLIANTGADTIRINDEEAPLWKPV
LITENSTVSFGCKGSRAYLAAAGIEVPAVMESKSTVVRGSIIGLHGRALQKDEL
NIGENSALSTLRLSLQAHVTIGTSDVVSQWPWEISVIEDESRHSAGGISICILH
SRINKAFSEAPRVTPQSDRMGYRLKGPELDLTAPLEWISAAYSAGAAOMPAGKPYI
PCLQRQDDWPAIRGIAHLISADLPIVSIQMPGEHVQFEPVSLQEAALAVREQHIIK
ELKTRKMKEWLIT"
58419..59159
/gene="ycso"
CDS 58419..59159
/gene="ycso"
/function="unknown"
/note="similar to transcriptional regulator (iclr family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12217.1"
/db_xref="GI:2632710"
/translation="MQNKNKTVMKSMALLNLNLFHLKPSLTLSELVSLTGMPKTSVHRMV

SSLEMGFLSRDASGAYSLGLVFLFEGQLVADRDLDIRKIAKPVMBEELREYDEAVOLLI
MRDNEAIYVEKITEGTOTVRLYTAIGRRSPLYAGACARSILSFLPREIEAYIKOTEL
ISTSGTITPEKILQIEDASVQNGYTVYSSELENYTAAGAPIFNHERQVAAGISIA
GFPAFTEDRLPYLITERKDAALQISRK"
59234..59899
/gene="ycsk"
CDS 59234..59875
/gene="ycsk"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12218.1"
/db_xref="GI:2632711"
/db_xref="SWISS-PROT:P42969"
/translation="MVLRYTALGDSLTTGRGSLGSPGFQVORFQGMMEADLKTRTAIN
IFARSLNTEILGLLSYPVYQKICRDADMTITTCGNDLIDSVLAYOTSKDEIFSR
VSAHCHEPFKMTAKVAEIKGENPSPAIVRNVLNPFPEIDTAKKITTSPNSHLETL
ASAPHVKIADAYSIFPKKEQEYLSFDGVHFNKSGYQMAEAHVHLGKELSVS"
59880..59899
/gene="ycsk"
terminator 60048..60293
/gene="yczi"
CDS 60048..60293
/gene="yczi"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12219.1"
/db_xref="GI:2632712"
/db_xref="SPTREMBL:O31483"
/translation="MFRIFKMSFAVILIIIALIAFNYTEHTSVIQSVMLVFLGAVMFM
QGLEERKENDGSGAFNIYAVFVMSVSLIGFTLHII"
complement(60299..60586)
/gene="yczj"
CDS complement(60299..60586)
/gene="yczj"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12220.1"
/db_xref="GI:2632713"
/db_xref="SPTREMBL:O31484"
/translation="MVRKAMHLIKGLEQEFEDAFRQAAPIITSGMKGYITHLSKCM
ETHKYLLLVETLEDTEGFRGSSEYQSWKALLHRFYTPFTVEHQDV"
60737..62743
/gene="pbpc"
CDS 60737..62743
/gene="pbpc"
/function="peptidoglycan biosynthesis"
/note="alternate gene name: ycsM, yzsa"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein 3"
/protein_id="CAB12221.1"
/db_xref="GI:2632714"
/db_xref="SWISS-PROT:P42971"
/translation="MLKKTLLVFLCVGLIGLIGCKTSDPSDRMEAFVQWQDQFDD
DMYQSTKDVKEISKKDFNRYKAIYEQAGVKNLKVTAAGEVDKDDQDNKTKHPIYK
VSNNTNAGVSKNTAVLKEKTDDESNINDWDPSIFKQLADDKTVQIMSEPRKG
QIYDRNGKLVANTDVEIGIVPELGDKKEVKIKELAKDLDEDDIKKKLDQGWYK
DDSFVFLKVKPKDQEKLVSEANTSLQGVTRTNVSSRYPYGKTAHLYGYVRAITAEEL
KKKKEGYSOTSNIGIAGLENVYEDKLRGTGMKIYVPQTGEVIAEKKAKDGEDHLT
IDIKTMKLYDELKDDSGAVALQPKGTETLALVSPASYPDNGFIKGSQKEMKLLNK
DKNPESAKFNKYAPGSTIKPIAAAIGIKNGTLKADEKKTIGKEWQKSSHGVS
TVSRERLQVDLENALITSDNIYFAQNALDMGADTETKLGTEFGFSEDVPEPIQKS
STANDKLSDDLIIADTGYGQGMQMSPLHATATYFVDNGDLVKPLTIKKDSOTADY
WIKQVYTKGCAADITKGLGVVEDEKRSAYQPVYKGTIVAGTKGTAEKLTSKDDKDG
ENGWFYGYDENKDLLVAMMIQNVDRGGSHYVVEKAKKQFQSN"
62845..63747
/gene="ycsn"
CDS 62845..63747
/gene="ycsn"

```

/function="unknown"
/notes="similar to aryl-alcohol dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB12222.1"
/db_xref="GI:2632715"
/db_xref="SWISS-PROT:P42972"
/translation="MQRIQLAEDLQFSRVIRHGLWRLNEWNSDAELLSLIEWCIDIHGI
TFTHADIVGYCTCKLFQNALALSPGURENIELVTKGIVLESPERPAHRSHHYNTS
KSHILASVEQSLMNLRTDYIDMLLTHRPDPLMDPEGAFAQALCKCKVRYGVSNF
KHQYRMLESYLPKLVNTOIELSAYELEMLDGLNLCQEKRIPPMAWSPLAGSKYF
TENDKDRVRITALESGEIGAASLDEVYAWLYTHPAGIMPVIGSGKRERISAAIN
ALSXYKLDQWFRIFTAVQGYDIP"
63933..66051
/gene="ydaa"
63933..66017
/gene="ydaa"
/function="unknown"
/notes="similar to transcriptional antiterminator (BglG
family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12223.1"
/db_xref="GI:2632716"
/db_xref="SPTREMBL:P96574"
/translation="MYMTAREQKLLKHLLONNRYITVTETLAELMQVSTRTIHRLEKSI
KPLMETVGLTDKQPGKGLKAVGSPGKOKLLTDLSEYQHEYSADERKLLILCSLLES
QEPVKLYTLANDLOVNTATVSVDLEKWLSPGGLTLIRKRGFGLIOLIGPENAKRKI
VGNLIVRLDQMFLEAVELNKKTKTSSRMFGVVGSGELLKMRILFQLKEKIAFS
LSDSYIALVHLTYAERIKLETITMEQNELEELMNAKEYSSALEIAGELERAFYV
TIPEAEVGYTHLRSANRKYETKAEIELETALQTKLIAFTISDKIRMDLTKNYS
LIEGLIAHLEPAVSRIKENIETYNPKQIKRDYFLLYMAIEEGVEKYFPGMSFSDDE
IAFIVLHGSALEIKKEAKVKALVVCSSGIGSSKMLASRLKPELPEIESFDMSSLIE
LKGDQVADYMTVSTPIPVENIDIVYSPILLNEEDANQVKQYIKRKIPLILNKRSS
KEEAQADVPDMLEAESIGRYMEVIOQVLRHFTLAQLKTNPDHSMLLLELFQQLKRD
GLIRPEKAAVCLAEREKQGGIGPTNMALYHLDNDEIVLPFFKMFDLSTPYEVDGM
DGLNLRMTIRILVMAFGSLAEGSEILSAISSAIIESGESMAGFOEGGQELYQRLNR
IPFTWMKEKNIL"
66028..66051
/gene="ydaa"
66228..67625
/gene="ydaa"
66228..67625
/gene="ydaa"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12224.1"
/db_xref="GI:2632717"
/db_xref="SPTREMBL:P96575"
/translation="METLQHLILHDMPSSEIEAVKSGDHTLTLYKGYRKRINOLANAM
LOKGIQKGRVALLCKNGHPASTVMAFALEIGAVVVPVSWQLKPYEMTGILKASEPKA
MFGYAEFELDEVLPELSSCVTMETATVETSAEFALFAGDHPHLPETEMVSPDDT
ALLMFTSTGNPKRCMITHGGIYRVKSSNSIAKGLRFLACHPIYHTSALICIM
LGTFAETTVFTKQDDPVHMLKVIIEERIQTVMALPVFTYILLEAWERHQDLSLVI
LMGTGKVPSSLSRYLDIGIPLAHGYGSTAWGISTWPDGMOKAASAGKPVAGVK
WKVEDPLTGEELPQGEIGEIVVHTPELFKGYEDNEPTAKVLQNGWFERTGDSGYDED
GFIFTGKVKVILYGGDNVYPDQVEEVIIQIPGILETAVVGIPDPLIGERPKAFIVK
NGGQRTIIEEDVIAFL"
complement(67760..68305)
/gene="ydaa"
complement(67760..68305)
/gene="ydaa"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12225.1"
/db_xref="GI:2632718"
/db_xref="SPTREMBL:P96576"
/translation="MIAGYIMAAENQTLNQWNTINQLGTRGDSILEVFGPGYCMQOM
LKREXDVHLHGIDVSEAMLKLAARVPKPKVRLIQGSIETETPLPASFYDKVIVSNNT

```

gene

CDS

terminator

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

terminator

gene

CDS

```

IWNQTKGIQIYRALKPGKAAITMQPREADASPEKTKSFGQRIADAFKAAGFEDID
IQFKNIKPELSVCATAKKPAT"
68513..69373
/gene="ydaa"
68513..69373
/gene="ydaa"
/function="unknown"
/notes="similar to alcohol dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB12226.1"
/db_xref="GI:2632719"
/db_xref="SWISS-PROT:P80873"
/translation="NANTPKELPAQTSQRPQGIEMSNPSPYVEYEDYKQADKLKGV
ALITGGDSIGRAVSAYAKAGADIAIVYKDEHDAEETKRVEQGVKCLLIAGDVG
EEFCNEAVEKTVKELGLDILVNNAGQHPKESIKDITSEOLHRTFKTFYSQFVLT
KKAIDYLPKGSAINFTSINPVYGNPTLIDYATKGANATFTMAQALVKGDIIRVNA
VAPGPIWTPLPATPPEETVAQGGQTPMGRPQPVHEVGCYVLLASDESSYWTQTL
HYNGNFTVT"
69389..69892
/gene="ydaa"
69389..69892
/gene="ydaa"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12227.1"
/db_xref="GI:2632720"
/db_xref="SPTREMBL:P96578"
/translation="MGITKEEVNSIYQAGIVLTDVEVDQIOLMDYGLGKERKVLQIL
FYVNTDQYCSKELVLPFCQTPCEHRHPVVDQEGKQETFCRCYKVIYLYVEGKTP
PKVLPPQDREHYTVWHEIELEPPGGQYTIPTPNTKHWFOAGEGAVVTMSSTSTDKHD
IFTDPR"
69977..70528
/gene="ydaa"
69977..70528
/gene="ydaa"
/function="unknown"
/notes="similar to acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAB12228.1"
/db_xref="GI:2632721"
/db_xref="SPTREMBL:P96579"
/translation="MFTCKVNEHITRLEPKDAERLAEILIQNOORLGKWLFFAENP
SSADTVRETIPDWRROYADLNGIEAGLLYDGLSGMISLHNLDOVNRKAEIGYNTAK
EFEGKIITACRKLITYAFELELNRVAlCAAVGNKSRAPVPRIGFLEEGKARDGL
YVNGMHDLVYYSLLKREWEK"
70606..71445
/gene="ydaa"
70606..71028
/gene="ydaa"
/function="unknown"
/notes="alternate gene name: yzza
similar to general stress protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12229.1"
/db_xref="GI:2632722"
/db_xref="SWISS-PROT:P80238"
/translation="MNQDQIKOKVLDVLDLHHKVGSLATVQKPKHSRYMTFFHDGLTI
YTPSKETHKAEIENPNPNHILGYDCFGDGYVEVAGAKINNNAELKDKIWNSSK
LERWFDGKDDPNLVILEIEPEDIRLNMAGEKTPVSLEL"
71424..71445
/gene="ydaa"
71534..72373
/gene="ydaa"
71534..72343
/gene="ydaa"
/function="unknown"
/notes="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11

```


VLTSELITXGSKDSAEITQIPAGASSGNQQLVIEYEAENLLISPSSLTVVIDEPV
KTLKLDGDSRRKTKVLNKLNSQSAQGYHNVSLKFGVMKGCVCVRQOTSGNWKIYPD
SRULTADSSAKGTSLDHPYPYPAQSGNTAEETAIVIPDNPSSAEIAAQCWKECYLT
VDSVSAYVTESELSKKIDKPTIVIGYDKHNGVKVLLKOAGLOAKGENLLAERVL
KAEGSQPVLPFAAASDAILTKKISVMTDQTYTQOLSGDILSLSKLOTESKESDKLT
LENFGAGDITIGADKTSAAHYFPASAVLDENQSAKLSLKLSKETIOASTAESAS
AOELKVMINGQPHSVDELGKEDKNGFYHVVVKVDPKLQKNRYIDIQFVTTGLKE
NPNCTTDEKWWFIDKNSTLSVAIKGSPSADQEWPLPYAGNQDQTTLIVLPDTVS
QSKLEELSLVTESEFNGEAQHSYTVKKSDDVTANDAKGRNLFICGINOFSLLKPKSSD
LLVPOEKNGSFDVSSFFEMLNETHKQVFTQASVMDSRVTAVEFAPSQGDGTAVTKEIL
SYLNSDESATVLNETNSQQVFTNHQOLKSETNSDAAEQPTQDHSQKWMYIGVLALIM
VVAVFVWIAVRKKRKTDTE"
82762. .82782
/gene="ydan"
83066. .83091
/gene="ydan"
83235. .85089
/gene="ydan"
83235. .85058
/gene="ydan"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12239.1"
/db_xref="GI:2632732"
/db_xref="SPTREMBL:P96589"
/translation="MYHSIKRFLGKPKLSQAAGEQKLTKLKALMLSSDALSSVAYG
TEQILILITASAAFWYSPIAVGVLLILLALLLSYRQIIYAYPOGGGAYIVSKENL
GEKFLIAGSLVDYIILTVNVSISAGTDAITSAPPAHDYHVPALFVLVIMILNL
RGLSESAILIAYPVYLFVWALLVLIAGVLFKLMGTQIQDOPAHSTSLGTPVAGITFLLL
LKAFSSCCSALTGYEASINAIKPNPARNARTLAMMGILLAILFSGITVLAIGYV
TAPKDDTVVSOIASETFGRNVFYVITQGVTSLLVLAANTGFSAFQPLAFNLARGYV
MPRFTVVRGRLGPNGLIFLGPAIVLILFSGQTEHLIPLXAVGVIEPTLSQTM
CMWKIKQPKPGWICKMLINSCGALISFWLSILFVTKENVVWVPLIFMPVIVLFFAI
KNHTVAGEQLRVIDKEPEEIKGTVIPVAGVTVTVOKSIHYAKLSUDQVIAVHVSF
DRQEKFEKRWEELNGVRLVTLHSSYRSLVHPDFKFELETVEAKAKKEQFSVMVLFP
QFITKKRWHTILHNQSAFLRLVFLRWFKKIMVATLPYHFKK"
85066. .85089
/gene="ydan"
85117. .85566
/gene="mutr"
85117. .85566
/gene="mutr"
/codon_start=1
/transl_table=11
/product="mutator protein"
/protein_id="CAB12240.1"
/db_xref="GI:2632733"
/db_xref="SPTREMBL:P96590"
/translation="MYTQGAFVIVLNESQOILLVYKRDVPLWDLUPGGRVDPGESAEBA
AVREILREETGYNAALSAGIGVYORPKFDDEQHLFEGSITGGOAMADGTETAGLKWVSP
GRLPFLWPNRKRQINDFNKGAQDVNTVVDKSLGLLAILDLKRLRGK"
85633. .87380
/gene="ydan"
85633. .87357
/gene="ydan"
/function="unknown"
/note="similar to pyruvate oxidase"
/codon_start=1
/transl_table=11
/protein_id="CAB12241.1"
/db_xref="GI:2632734"
/db_xref="SPTREMBL:P96591"
/translation="MAHKTAGQAMTELLEQWGVHDVYVIGPQDSINERIEELRHERNOL
KFTOTRHEEVAALAAAEAAKLTGKIGVCLSTAGPAGVHLLNGLYDAKADAPVIAIAG
VSSGVEGRDYFOEIKLEQFEDVAVFNREHVSAESLPDLINQAIRTAIRYAKKGAVILS
VSDDLFAEKTREPVYTSPIVIEGNLPPKKEOLVTCQYITNNAKKPIILLAGQGMKKAK
RELLEFADKAAPIVVTLPAKGVVPDKHPHFLGNLQIGTKXPAYEAMEECDLLIMGT
SPFYRDLPDPTAIQLDSDPAKIGKRPVYAGLVCDLSALGRELTYETIERKEDRRFL
NACTHEQHWHNWELEKDETEATPLKPPQVVARLQEAADDAVLSDVGTVMVMARH
FKMANODFTVSSWATMGGCLFGAIAASISEPERQAIYAVCGDGFSSWVMDLPTAVK
YKLPITVILNLENLGMIEVQQVKGNIDVYVTKLQNVDYAAFAESCCAGKIGKVTAAE

LAPAFHEALHSDQPVVVDVMIGNBPPLPGKITQYQAKGFSKYMULKNFENQKFEMPSP
KKSLLRLF"
87366. .87380
/gene="ydan"
87534. .87824
/gene="ydan"
87534. .87824
/gene="ydan"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12242.1"
/db_xref="GI:2632735"
/db_xref="SPTREMBL:P96592"
/translation="MLILIKSYRTLIIIGYSEFFSPKKKGKAEPTKKVPKNKMPPIN
TWKEVCPAVNAVIREMILSTIMAMPGATLIIFPLDLWAATEHKMPITIGTIHK"
complement(87926. .87942)
/gene="ydan"
complement(87926. .89225)
/gene="ydan"
complement(87948. .89225)
/gene="ydan"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12243.1"
/db_xref="GI:2632736"
/db_xref="SWISS-PROT:P96593"
/translation="MMNKDITAQSPRSKAVODALDGKTRGPRGLLPFLGPAFIAIAY
IDPQNFATNISAGSKYGYMLLWILFNSIMALLIQSLSAKGIATGKNLPEVAREEFP
KPVSTIGLWIOGELVITADLAEFICAGALGYLLFGIPMLEASIIAAIGSFALILEQRR
KYRSLEAGIAGMLFVVITAFALOTFEAKPDVAVSMKGLFVPAPFHTGDSVLLAAGITLGA
TVMPEHAIYLSALTQRRVVGKTDSPERKKIFRFEEDILILAMLIAGAINASMLIVAAAL
FFKNGLFVEDDQVAFQQFGHLVSPMSALFGILLVAGLSSSSVGTLSGDMVMOGFIN
YRIPLYNARWITTIIPPLIIASGVNPTTALVLSQVLSVFGTAFALIPILMFTSNKRIM
GSLINAKWITVYSWLIIVIALVINALNVFLIVDFR"
complement(89411. .89437)
/gene="ydas"
complement(89411. .89712)
/gene="ydas"
complement(89455. .89712)
/gene="ydas"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12244.1"
/db_xref="GI:2632737"
/db_xref="SPTREMBL:P96594"
/translation="MLSFLVSLVVAIVIGLIGSAIVGNRLPGGIFGSMIAGLIGAWIG
HGLLGTWGPSIAGAPFAIPAIIGAAIFVLLGLIFRGLRKEA"
complement(89790. .90242)
/gene="ydan"
complement(89790. .90242)
/gene="ydan"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12245.1"
/db_xref="GI:2632738"
/db_xref="SPTREMBL:P96595"
/translation="MPWSMKDYPASLKNLEKPVKKAIIDIANAMIDEGYEESRAIPTA
TSKAKWAENASTDEIDDFLTHDDETERDADPSSGSGPELMNKAHVYIKKKHGWAKYA
EGAKRVSEIKDTKKKEATERAKEIAAHKGTEVIVHLADGSVQRKIKTGS"
90360. .91210
/gene="ydba"
90360. .91178
/gene="ydba"
/function="unknown"
/note="similar to ABC transporter (binding protein)"
/codon_start=1
/transl_table=11

```
/protein_id="CAB12246.1"
/db_xref="GI:2632739"
/db_xref="SPTREMBL:P96596"
/translation="MPTEKEDLHQOLKEIEKWEKDOOKVWFWEKLSRLPFLQDLKLT
PEFTQIKIGLLDEVGFVGTGOYLISEKQIIRMFQKLLPEEIPESLEDIRKAPLPV
MDETAEGMGKRTNAATVQAGATTGGVFTLAADIPAVLGSLTLQLOIAVAYGDPK
EKKEEYFIVKQLQTSADVVGKRSILQELKHYYDODRTYKNVASIQGWREVVLGYROT
FGWKKLFQIVPVAQMGVGAANRSTLNDITETGMMLYKKRILRLKETEREME"
91185..91210
/gene="ydbA"
/gene="ydbA"
91307..91741
/gene="gsIB"
91307..91678
/gene="gsIB"
/codon_start=1
/transl_table=1
/product="general stress protein"
/protein_id="CAB12247.1"
/db_xref="GI:2632740"
/db_xref="SWISS-PROT:P26907"
/translation="MADNNKMSREEAGRGKGGETTSKNHDKFEYQIEGKGGEATSKNH
DKFEYQIEGKGGEATSKNHDKFEYQIEGKGGEATSENHDKFEYQIEGRKGGEATSK
NHDKFEYQIEGSGGNARNND"
91716..91741
/gene="gsIB"
91810..92151
/gene="ydbB"
91810..92151
/gene="ydbB"
/function="unknown"
/codon_start=1
/transl_table=1
/protein_id="CAB12248.1"
/db_xref="GI:2632741"
/db_xref="SPTREMBL:P96597"
/translation="MSKQNLAIANEVKEKHANFSIYENDHCVRLLAVTIGEYDWHH
HPSDDELIVLEGEILLDFDKKETAVLKANDSLIPKGTVHRTSYVTVNLCVEHKO
AEYVILIEQPC"
92145..92504
/gene="ydbC"
92145..92504
/gene="ydbC"
/function="unknown"
/codon_start=1
/transl_table=1
/protein_id="CAB12249.1"
/db_xref="GI:2632742"
/db_xref="SPTREMBL:P96598"
/translation="MLIKKMICEDANAFAKAQSOWEALSHISGFYKQAGGWRKT
IDPLTAELISVWENREAYDHFENEHDSIYEENDOKAVILSIETVYVEEDKPFVHDL
LHNPDIRYEPNWTVLKA"
complement(92541..93362)
/gene="ydbD"
complement(92541..93362)
/gene="ydbD"
/function="unknown"
/notes="similar to manganese-containing catalase"
/codon_start=1
/transl_table=1
/protein_id="CAB12250.1"
/db_xref="GI:2632743"
/db_xref="SWISS-PROT:P98078"
/translation="MFKHTKMLQHPAKPDRPDFLFAKKMOEILGQGFETISVAMQYLF
QGNTRGNEYKDLMDTATEELGHVEMTATMIARLEADPLDQQAADDPVIGSIL
GGMNPHAIYVSLGAMPESSTGVPSGGYIVASGNLLADFRANLNAESQGRLOVARLF
EMTDDKGVKMDLSELLARDTMHONOLAAIKLEAEOEGPVVPCTFPKALEKQESHOL
INFSEGEVSAEONNLNEKAPDGEAFYEVREAKTFGEKPELKAPPFPVHNTLPGRE"
complement(93453..94505)
/gene="ydbE"
complement(93453..94505)
/gene="ydbE"
/function="unknown"
/notes="similar to C4-dicarboxylate binding protein"
```

```
/codon_start=1
/transl_table=11
/protein_id="CAB12251.1"
/db_xref="GI:2632744"
/db_xref="SPTREMBL:P96600"
/translation="MKSLLACLALMIAGIATALFTGFHDHTGNKKIVYDDQOEGLOQDQ
IVFEGSHVAENTPKGLAANKFADLVNEKSGGKIKIEVFPNGSLYSDIEEALQNGLA
VQFTAPSTSKGLMSPGWLDPYAFYFDYNAVKKGLNGSIGTQIFDLSLKNQLKGLA
YNTNGFKQITTNQGVKTPDQLRIMQSDVIEDQFKLLGATPHQOESNTPFQLL
ENNVVDGEENTISNYSKKFYNVODYLTISSHGYLGAVMTDEHFMAKQTPETREILT
EAMKETTEWNETYAEQNNKEQLEIKKNSAHHIELSDKEKEQWEMKRLDPVYRQYEPI
FGRLIRELLELRKDS"
94575..96182
/gene="ydbF"
94575..96182
/gene="ydbF"
/function="unknown"
/notes="similar to two-component sensor histidine kinase
[ydbG]"
/codon_start=1
/transl_table=11
/protein_id="CAB12252.1"
/db_xref="GI:2632745"
/db_xref="SPTREMBL:P96601"
/translation="MNNKKLSIRWKITILSYILVIFSLIGGIVLIGNIOHTEREELK
KRLMNTARTVSEMTVEVKEALARKKQTEAVRHAVEEIRMEINADYIVVMNMHRIYTHP
VSTSIGKSPGADBEAFAEHIYFSEAKGEIGTAVRAFYPVKDQDLQIGVVLVYKTL
PGIADILLHLKRDIAFIWLTGLPGLAGSLLARHIKKOMFOLSPHEIVRMVYERTAT
FHSNEGVIADNRLVITIFNEKAKQIFEVQGDLLKVIWEVLKDSRLPELVERNKAV
YNEEIRVSGKVMSSRIPIVMKKKVIQAVAFQDRTAAKMAEELTVGRNFEVREALRVQ
NHEHMKLHTIAGLIQKSEKALQAFQASTEQENYTEFLHRSIQNDAAAGLLLSKI
RRGRELGAIVHIDENSLLQFQPEHVDQHDIVVLIENLIENAFGSEFVQSBDKRIDIS
IEQTDILAILIEDNGGCIPTHPRLYDKGFTVNTGGTGYGLYLVKQIIDKSGGTI
EVDSHAGQTSFSIVFPMKGEEAQHGS"
96172..96852
/gene="ydbG"
96172..96852
/gene="ydbG"
/function="unknown"
/notes="similar to two-component response regulator [ydbF]"
/codon_start=1
/transl_table=11
/protein_id="CAB12253.1"
/db_xref="GI:2632746"
/db_xref="SPTREMBL:P96602"
/translation="MARKEMKVLIEDDPMVQEVNKKDFITTVKGVTVCATAGNGEEM
KLIKEQPDVLIVLDVYMKKDKIKTLQEIQRKQLEVDIVVYSAKDKETISLMLQNGA
VDILKPFKLRMRQALEKYQYKQKTEANDTLSOEQLDALNIPQQAQVQDLPLKGLNH
FTWNEVTAFLKQQTASUSAEVAKALGIARTYARRYLDYLEKTGIILKLDVQYGVGVRP
VNRVYLVKG"
96973..98264
/gene="ydbH"
96973..98238
/gene="ydbH"
/function="unknown"
/notes="similar to C4-dicarboxylate transport protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12254.1"
/db_xref="GI:2632747"
/db_xref="SPTREMBL:P96603"
/translation="MKFLNLTQVITAVIIGVIVGLVWPDPVGMKEMKPLGDTFINAVK
MVTAPIFFIVILGIAKMDMKVKVGGKAFIYFEVVTTLALTIGLVVNMKPGAG
LDYSLEKGSVOVTOGGQIDWIEFTHIVPSNMVDAFAKGDILQVLPFSILFGVG
LAALGKSGKSIDFFDKVSHVEFKIGYIMRAAPIGAFGAMAYTIGHGLDSIKPLAS
LMSVYITMFLFVFNALNIICKLYGFSLWNLRFRIKDELLIIVLGTSSESIVLPRMDK
MERIGCSKVVGVIPIPGYSFNLDGTSIYLSMATVFLAQVGVGVDLSIGQOITILVLM
VTLKGAAGVTGSGFIVLASTLSALQVIPLEGLALLGLGDRFMSSEGRAIVNLIGNGIAT
IIVAKSENEFEDEAKIEAVEGMMKMKMTAV"
98247..98264
/gene="ydbH"
98386..99682
/gene="ydbI"
terminator
gene
```

```
CDS
98386..99438
/gene="ydbI"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12255.1"
/db_xref="GI:2632748"
/db_xref="SPTREMBL:P96604"
/translation="MAALVKFFQHPCGVRFRSFVVLVTGLYLKFSKMINLILITFTFTE
LMRLVLVVRQVSOFFRVSRQVWITFLYMLLAVLLTVGSGFYFYPVAAQIQOLVKQI
KHIAHPDSIPFDEITSVFGDINISSYKGFENVVYTLADISFGLQVAAQILISM
FFLEKKRUSEEMAKFTSKLRVFEETAFPGSKFARTFGKVALEAQFIIVWNCILTF
IALWIMHFPQLGSLIMVFFLGLIPVAGVISLIPLSIIAYSTGGGMYIYIVLVIFA
IHAIETFLNPKLMSAKTELPITFTTTLIFSEHFGLWGLIIGIPFVFLDLIDVT
NKEDRSKGPRENSDKK"
99661..99682
/gene="ydbI"
99715..100641
/gene="ydbJ"
99715..100641
/gene="ydbJ"
/function="unknown"
/notes="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB12256.1"
/db_xref="GI:2632749"
/db_xref="SPTREMBL:P96605"
/translation="MAEPILHIEGLDKKIGSKQILKQISMVWMEGIIIGLPGNSGK
TTLRIIVGLLKONSGSVTISGFQHDTEFEKAMEAVGAIVENPEFYPLTGWENLKHF
ANMHRKIADELRDEYVERVGLTSAIHDKVKYTSLGMQRQLGTAQAILHRPKLILIDEP
TNGLPDAGKDFROHKELAEMEGTAVLPATHLLREVEDLCDRVIIIOKGEIKAEVSL
QCTDOTEKAIIEVOPOEKALNWLTCNOYQAESQDGTIVVEVAKENIPELNSLVGQD
LNVFSITPTQSLDEEFKATTAHQEGEGLY"
100661..101401
/gene="ydbK"
100661..101401
/gene="ydbK"
/function="unknown"
/notes="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12257.1"
/db_xref="GI:2632750"
/db_xref="SPTREMBL:P96606"
/translation="MKLFNRKVTLSLILMAVQFFMALIKRIVISAGTDENFTGYL
STPSSLNILLQALTIIVAAATIVSMEDFKKTIKPELLIRPVKQKVFVSKLITVVMVSFY
LYLAYIILALFGLLFFGTSVTAESKTLVNTLALIGSNWLEAVMMGLFGLCSSLFR
NSAVAVSVSVLYGASTLVQLMKLFENKWSGFLFLFANDTFTQYRSGETALFSGMTPL
FSIGILIHAIFFIYVVGWMCCKRDVRV"
101496..101859
/gene="ydbL"
101496..101831
/gene="ydbL"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12258.1"
/db_xref="GI:2632751"
/db_xref="SPTREMBL:P96607"
/translation="WRNFITALPILVLLGLFSFVSMFQFEHLVYFRILALGLFSLVGLY
MIYKMKTGIRYFTIYLYASWIVLAATAPEEPIFSFFFGGLVMTNGILTYMLIILGM
KQDRDANPV"
101833..101859
/gene="ydbL"
101959..103125
/gene="ydbM"
101959..103104
/gene="ydbM"
/function="unknown"
/notes="similar to butyryl-CoA dehydrogenase"
/codon_start=1
```

```
/transl_table=11
/protein_id="CAB12259.1"
/db_xref="GI:2632752"
/db_xref="SPTREMBL:P96608"
/translation="MSLFIQNDQORQWMEKIGRIADEFOQTAAEDDEQGRFPAEKIQK
LRDAGYTAUTLPASHGGGSGISYVDMLLFQERLARGDAPTALSIGWHLISVIGELGEGNS
WDEDFAFVAKEVQNGAVINRAATEAKTSGPTRGGRPGTHAVKDKGKWA/VNGRKTFTT
MSQALDYFLVTAWIEDKQTVTELIIHKDDPGLSIEETWDMAMRATGSHDLVINEVML
DENKVELLQGGPRGAKPNGWLHPALYILGVAQAARDYAVQFASEYSPNSLNGPIKRV
PAVOQTGMELELLNARHFLHIAOLYDDPVRRPHLTSELGAAKHTVTNAALSVDVK
AMRIVCAKSLERTNPLQRYRDVRAGLHNPMPDDAVIHKLAEEAFES"
103111..103125
/gene="ydbM"
complement(103129..103308)
/gene="ydbN"
complement(103129..103308)
/gene="ydbN"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12260.1"
/db_xref="GI:2632753"
/db_xref="SPTREMBL:P96609"
/translation="MNTNMSGRRSMPKRKVTKYQQLIQENKEAIMGNPKLMNVIIDR
IDRKHKNLQEQNNT"
103673..104545
/gene="ydbo"
103673..104545
/gene="ydbo"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12261.1"
/db_xref="GI:2632754"
/db_xref="SPTREMBL:P96610"
/translation="MERTENLKKRGKALNIFAYVILAVKVLIGTLYHSEALRADG
GKAIGAGTTETPNLIAAWTALGSAVFMVGIYLYNKRLAASIKKSALMAAAKDSRSDAF
VSAGATIGVFSQSLKLPWVDPTAFIIGIICCTAMDIFKADASHLTDGFHLKDLPEY
KQTVGRIENVHRUKDKARYLGSVTHIEMVITVDPKLITVEEGHGVADVEDKIKHEHD
VTHVHVHPEDDIK"
complement(104489..104516)
/gene="ydbP"
complement(104489..104516)
/gene="ydbP"
complement(104560..104880)
/gene="ydbP"
complement(104560..104880)
/gene="ydbP"
/function="unknown"
/notes="similar to thioredoxin"
/codon_start=1
/transl_table=11
/protein_id="CAB12262.1"
/db_xref="GI:2632755"
/db_xref="SPTREMBL:P96611"
/translation="MKKITTNEQFNELIQSDKEIIVKFYADWCPCDCTRMNMFIGDILIE
EYNQNDWYELUNKDELFDLAEKIQVMGIPSLLIIFNKEKTAHLHSANAKTPEEVTEFLIS
EHIS"
105055..106119
/gene="ddla"
105055..106119
/gene="ddla"
/EC_number="6.3.2.4"
/function="peptidoglycan biosynthesis"
/codon_start=1
/transl_table=11
/product="D-alanyl-D-alanine ligase A"
/protein_id="CAB12263.1"
/db_xref="GI:2632756"
/db_xref="SWISS-PROT:P96612"
/translation="MYGKSAEHNVSILQATALVIKALNTEKFDIHPIYITEKGWVRG
```

ARLTPVSNVKMLQEQGSAFSPAALNQBMFPOEASQONEADIDVFPPLHLGPNGEDG
TIOGMLLELLNPYGVNGVLASSAGMDKVMKHLEAQLVGLAQAKYAALFKDWTRSPKE
SCREVEQLGCPGVKPANLGSVSGKCRNRELQAKFELAQYDRKYVVEEINGR
EIEIGVLGNDPKPCSVVGEIAPKTDFYKAKYEDGDTDLMIPIAVYDEYATISDMA
IKAFKAIQDGLVRADFFLTPADGEVLINVENTMPGETPFPSMFFPLLMKEAGVEYADLIE
OLVELAKERHAERKOLIKHTE"
/db_xref="GI:107564
/gene="murF"
/EC_number="1.07564
/function="peptidoglycan biosynthesis"
/note="alternate gene name: ydbQ"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-
diaminopelate-D-alanyl-D-alanyl ligase"
/protein_id="CAB12264.1"
/db_xref="GI:2632757"
/db_xref="SWISS-PROT:P96613"
/translation="MIKRVKNIAMVKGTLANPOVEQTVIHGVADTDTRKLEOHQLFI
PLKGRFDGSEVEQAEAGAAVWMDRSPENHAVILYDDTLTAQOOLAKAYIQ
ELGTRVIGVTSNGKTTKMDIHAVLGTYRVKHTGNFNHIGLPLTLVLAMPENTEI
AVLEMRGAKGEIDLLRLANDPAVITNIGESHQDLGSRGIAKLEIINGLKED
GVLIYIGDEPLLQAYSCOTKYGTGTHNDYQLODYQSBEETHFTIKGIENTFTPI
LGKHNVMNAMAATAGAYFGIAPEDAAGLGLKVTGMRLLELTKDSGLSINDAYNA
SPTSKAAIQLTESLEGYCKKMLVLGDMLELGLDLETFHKECAVISPDKIDRVFTYG
KIGAFATGALKHFEKDRVSHYTEKKDLDLQAVKENASKGDLILFKASRGMKLEEIVRD
LIESPLS"
/db_xref="GI:109486
/gene="ydbR"
/EC_number="109448
/function="unknown"
/note="similar to ATP-dependent RNA helicase"
/codon_start=1
/transl_table=11
/protein_id="CAB12265.1"
/db_xref="GI:2632758"
/db_xref="SPTREMBL:P96614"
/translation="MVNHDITAIRRSNNITITFQDFNLSSDLKMAKAINRMGFEBAT
PIQATPIGLSKNDKQVIGQAGTGTGTAAGFGLVKEKINPESPNQIAIVIAPTRELAI
QVSEELYIKGDKRAKVLPIYGQDIDGRIQIRALKKNPNIIVGTPGLLDHINRRTIRL
NNVTVWDEADENLMGFIIDDIESILSNVSEHOTLLPSATMPAPIKRIAREWTEP
EHVKVKEMTVSNIQOEVLEQERKEDTLRLDLSQPELAIIVGTRKRVRDELAE
ALNLGCAEGLHGLTQAKRWALRKEGALVELVLDVAARDLIDISGTVHVNFD
VQDPESVHRIGRTGRAGTKGMATFFIPPREKSMRLAEQTTKRKMDRMKEPTLDEA
LTDEAPMYSKRYKNQSSKRRDGGGVRGGKGNNSRSDYKRSNDRSSGDRRQK
KSY"
/db_xref="GI:109464
/gene="ydbR"
/EC_number="109621
/gene="ydbS"
/db_xref="GI:2632759"
/gene="ydbS"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12266.1"
/db_xref="GI:2632759"
/db_xref="SPTREMBL:P96615"
/translation="MREQPKNQISPDGLKVRFLQELIIISAVCLLIVIAVAVLSYFHH
PYWISGLGAVWLGLSVTFIIPKVRHKVVRVEVHEHEIDIQSGIFVVRVIVPMVR
VQHVDTSGPLKKYNLATVKISTAATVHSIPALEMEEAADRLRDSISRLARVTDVV"
/db_xref="GI:110090
/gene="ydbT"
/EC_number="111571
/gene="ydbT"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12267.1"

/db_xref="GI:2632760"
/db_xref="SPTREMBL:P96616"
/translation="MMSEPKRLHPVAVILNCHTIIQIKNLIIPFFFFVYIVNSNHTV
RFGAIALGVLFILWVAASIIKRRFTYRDEDEFRIEGLFTVKRKYISIDRIQTMN
TSAGLQOQIFKLVLQIETAGGKEAEAVLSAIVEEAERKEAVFKKKAQRRENELO
EERLEAEEDLPSVEVQEHYRMNAKELLMAASTSGGIGVLIISAVFALLSQDDEVLPMD
WLPKFSFLQHASIGIYAVLIFIGLFIAMFISIAGMMFRYANFOIIRKEQELVIRGI
IEKHQVTIPLRQIAIKIKENIRQLFGFVTVSIVSAGGDRKEEGALTTILFPMIHK
KKLPHMLRTFTPEYTLBENCRRRLKRYLFRSFLFIPLCLIFFQPMGVLVS
ILLPIELFGYLAYKEAAMTNGDRLOLTSREIGRTAIVLKKRMQVCKFSQSFOKK
GRLYTISTSVKSSHMBELTVRDVGEEDAARILKWSYEKADG"
/complement(111815..111836)
/gene="ydcA"
/complement(111815..112422)
/gene="ydcA"
/complement(111823..112422)
/gene="ydcA"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12268.1"
/db_xref="GI:2632761"
/db_xref="SPTREMBL:P96617"
/translation="MFTRENFTQIFRLYPVVTILALQAVLWLFSLPAHSVVLWRD
TVGINLGVANGEMWRLITPILLHAGFTHLLFNMSIFLEAPALERMLGKARFLLYVA
GSGIIGNIGYVTEPLDYVHVAGSAGALFGLGVYLFNVLFRNELIGOEHSKMIITLLA
FAYLMSFINSNINMAHLFGLCGGFLSLFCVQKKERRY"
/db_xref="GI:112517
/gene="ydcB"
/EC_number="112517
/gene="ydcB"
/function="unknown"
/note="similar to holo- acyl-carrier protein synthase"
/codon_start=1
/transl_table=11
/protein_id="CAB12269.1"
/db_xref="GI:2632762"
/db_xref="SPTREMBL:P96618"
/translation="MIYGIGLIDITELKRIASMAGRKQFAERILTRSELDOYVELSEK
RKNEFLAGFAAKEAFSKAFGTGIGRLSQDIEIRKDQNGKPYICTKLSQAAVHVS
IHTHKVAAQVVERLSS"
/db_xref="GI:112943
/gene="ydcC"
/EC_number="112943
/gene="ydcC"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12270.1"
/db_xref="GI:2632763"
/db_xref="SPTREMBL:P96619"
/translation="MRLWVASVLPKRVPSPVLPKSNKRLPFTSIEKKGKLVKRSFV
LILLGLLAVLILSACGKTQODIVAGLDEKAKETSYKAKAMTETGSPQYVNEI
WHKKSLYRVYLENPKDNQVILNRNENGVFVLTSLNKSFRQSDWPNSSQVYVFE
SLVKQVQNSDAVFTAKERKYPFETKNYOHNMKLPTQETITFNKMDSPSSKVMMDTD
RKVMVKEFSSEFNKQFDKESFDEKKNMTLSQMDVATSAKPSDTPAVKTPLEPLGV
KLLEEKDIDSTEDGKRIIMTYGEGKSFLLIQEAKQIAKASSVTLNGEPVNLGVTIGAL
SDASLSWTYDGVLYLLSSKDLSEEMVTAKSMQGOSSK"
/db_xref="GI:114073
/gene="ydcC"
/EC_number="114179
/gene="dal"
/db_xref="GI:115348
/gene="dal"
/BC_number="5.1.1.1"
/note="alternate gene name: alr"
/codon_start=1
/transl_table=11
/product="D-alanine racemase"
/protein_id="CAB12271.1"
/db_xref="GI:2632764"
/db_xref="SWISS-PROT:P10725"

terminator

gene

CDS

```

/translacion="MSTKPFYRDTWAEIDLSAIKENSVNMKKKHIGEHVHLMVVKANA
YGHGDAETKAALDAGASCLAVAILDEAISLRKKGLKAPILVLGAVPPEYVIAAEYD
VLTGYSVEAQEAARHTKKGSLFHKLVDTCMNRNLGVKTEEVONVMAILDRNPRLK
CKGFTTFATADEKERGVFLMQFERFKELIAPLPKLNLMVHCANSAAGLRKKGFFNA
VRFGWGLRPSADMSDEIPOLRPAFTLHSTLSHVYKLRKGSYSVGAETATKEDT
WIGVPVGYACWLKRLKGTDLILVGRKLRKTAGRICQDMQDFWELDOETPPGKTVLLIG
RQGEYISMEIAGRLETINIEVACTISSRVPFRPLENGSI MEVRNPLQLQVNIS"
115464. .115745
/genes="ydcD"
CDS
115464. .115745
/genes="ydcD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12272.1"
/db_xref="GI:2632765"
/db_xref="SPTREMBL:P96621"
/translacion="MSESSARTEMKISLPENLVAELDGVAMREKRSRNELISQAVRAY
VSETRTHNRDLMRGYMEMAKINLTISSEAHFAECEATTVERLVSGG"
115750. .116131
/genes="ydcE"
CDS
115750. .116100
/genes="ydcE"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12273.1"
/db_xref="GI:2632766"
/db_xref="SPTREMBL:P96622"
/translacion="MIVKRGDVFADLSPVVGSEGGVVRPVLVIONDIGNRFSPAIY
AAITAQIQAKLPTHVEIDAKRYGFERDSVILLEQIRITDKORLTKITHLDDENMDK
VDEALQISLALIDF"
116113. .116131
/genes="ydcE"
terminator
116215. .117039
/genes="rsbR"
CDS
116215. .117039
/genes="rsbR"
/function="positive regulation of sigma-B activity in
response to salt and heat stress"
/notes="alternate gene name: ycxR"
/codon_start=1
/transl_table=11
/protein_id="CAB12274.1"
/db_xref="GI:2632767"
/db_xref="SWISS-PROT:P42409"
/translacion="MMSGNTVYQIAENQNELQLWTDTLKELSEQESYQLTDQYEN
ISKYIDILLSVKDENAESQISELARAQVQIGLSMKFLATALAEFWKRLYTKMNDK
RLPQESTELLQIDRFPSINTFQYSISWEKTVSLQKIALQELSAPLIPFVENI
TVMPLVGTIDFERAKRIMENLLNGVWRHSOVVLIDITGVPVDTMVAHHIIQASEAV
RLVGAKCLLAGIRPEIAQTIYNLGDLSQVITKNTLQKIQTALEMTDRKIVSLGE"
117044. .117409
/genes="rsbS"
CDS
117044. .117409
/genes="rsbS"
/function="negative regulation of sigma-B activity"
/notes="alternate gene name: ycxS"
/codon_start=1
/transl_table=11
/product="antagonist of RsbT (through dephosphorylation by
RsbX and activation)"
/protein_id="CAB12275.1"
/db_xref="GI:2632768"
/db_xref="SWISS-PROT:P42410"
/translacion="MRHPKIPILKNCLLVSIQWELDDQFALTQEDLLNKIYETGA
NGVVIDLTSVMDISFTAKVLGDVITMSKMGAKVLTGIQPAVAVATLIELGIALEEI
ETALDLEQLETLKRELGE"
117413. .117814
/genes="rsbT"
CDS
117413. .117814
/genes="rsbT"
/function="positive regulation of sigma-B activity"

```

```

/notes="alternate gene name: ycxT"
/codon_start=1
/transl_table=11
/product="switch protein/serine kinase (phosphorylation of
RsbS)"
/protein_id="CAB12276.1"
/db_xref="GI:2632769"
/translacion="MNDQSCVRIMTEWDIVARQLGRNVAKELGFGTVDQARITTAIS
ELARNIYLAGQSIGIEQVADRGKKGLKIIAEDQGPIDIRKVMEDGFSTSGGLGA
GLPGVKRLMDEFSLNSVAGETEIQAIKWLK"
117826. .118833
/genes="rsbU"
CDS
117826. .118833
/genes="rsbU"
/function="indirect positive regulation necessary for the
RsbX-dependent regulation of sigma-B activity"
/codon_start=1
/transl_table=11
/product="serine phosphatase (dephosphorylation of RsbV)"
/protein_id="CAB12277.1"
/db_xref="GI:2632770"
/db_xref="SWISS-PROT:P40399"
/translacion="MDFREIVQRTHQLLSRYIAELTETSLYQAOKFSRKTIEHQIIPP
EEIISHRKVLKELYPSEIDVFLHSLDFLIEVMIGYMAOEHQTLRGIQEIKSEIE
IAANYQOTLLGTQVPEEALDIGAISVPAKQMSGDYHYFKDKESINTAIADVIKGI
PAALCMSIKVAMDSLPETGHPSOVLKLNLRVQNVQDASMFITMFYANYNMDKHQF
TVASAGHEPGFYYSOKDNTFYDLPEAKGLVLTGSQDYDKQFDOHLEKGDMLVLFSDGV
TECRTENGFLERPDLQKLIIEBHMCSQAEMKNIYDLSLLKLDQFQLHDDFTLIVLRRK
V"
118895. .119224
/genes="rsbV"
CDS
118895. .119224
/genes="rsbV"
/function="positive regulation of sigma-B-dependent gene
expression"
/codon_start=1
/transl_table=11
/product="anti-anti-sigma factor (antagonist of RsbW)"
/protein_id="CAB12278.1"
/db_xref="GI:2632771"
/db_xref="SWISS-PROT:P17903"
/translacion="MNIINVQNQENDIQVNTAGEIDVYAPVLRKLVPLAEQAGADL
RICLDKVSYMDSTGLGVFVGTFKMKYKQGSGLKENLSERLIRLFDITGLKDIIDISA
KSEGGVQ"
119221. .119703
/genes="rsbW"
CDS
119221. .119703
/genes="rsbW"
/function="negative regulation of sigma-B-dependent gene
expression; phosphorylation of RsbY"
/codon_start=1
/transl_table=11
/product="Switch protein/serine kinase and anti-sigma
factor (inhibitory sigma-B binding protein)"
/protein_id="CAB12279.1"
/db_xref="GI:2632772"
/db_xref="SWISS-PROT:P17904"
/translacion="MKNNADYIEMKVPQAEVVGIIRTLSCGVASRMGYTYVDEIEDLK
IAYSEACTNVAQHAYKEBKNGEVSIRFGVFEDRLEVIVADEGDSDFDQKQODLGPYT
PSHTVVDQSEGGGLGLYMETLMDDEVVRQNVSGVTVMKTYLNGERVDHDDTTIKNYETN
"
119663. .120457
/genes="sigB"
CDS
119663. .120457
/genes="sigB"
/function="general stress sigma factor (class II genes)"
/notes="alternate gene name: rpoP"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-37 factor (sigma-B)"
/protein_id="CAB12280.1"
/db_xref="GI:2632773"
/translacion="MIMTQPSKTTKLTKEVDRLISDYQTKQDEQAQETLVRVVTNLV

```

```

DMLAKKYSKGSKSPHEDLROVMGIMLLGAIKRYPDPVVGKSPFAPIPTILIGEIKRFLRD
KWSVHVPRRIKELGPRIKMAVDQLTETQSRPKVEIEAFLDVSEEVLETHWEMGKS
YQALSDHSLEADSDGTSVTILDIQSGDEGYSERVNQMLQSVLHVLSDRKQIIDL
TYIQNSQMETGILGISQMHVSRLQRKAVKRLREALIEDPSMELM"
join(120457. .121056,121070. .121089)
/gene="rsbx"
CDS
120457. .121056
/gene="rsbx"
/function="indirect negative regulation of
sigma-B-dependent gene expression"
/codon_start=1
/transl_table=11
/product="serine phosphatase (dephosphorylation of rsbs)"
/protein_id="CAB12281.1"
/db_xref="GI:2632778"
/db_xref="SWISS-PROT:PI7906"
/translacion="MIQVEENEHIQTLVYQLNKEGKSIGDSFFMKADKRELICAVAD
TGCGLANESSAAIKDLVENYASEDVESIERNQAMKNKGATASILKINFEQRF
TVCVGNVRFILHSPSGESFYPLPSIGYLSGPKQKYKTHATYEGKSKFIHTDGLNV
PIDRSHLKGQSVEEISNSLKWYTSRKDDLIYLQLS"
121070. .121089
/gene="rsbx"
terminator
121299. .121592
/gene="ydcf"
CDS
121299. .121592
/gene="ydcf"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12282.1"
/db_xref="GI:2632775"
/db_xref="SPTREMBL:P96623"
/translacion="MERKEHCFCCKQFSYGLSMRNIGRFLSWSTCPSHNVIMEFFH
VLQKQEDLCITGIVLEGAEAKWGEIITGELSKDMPATAVSRVLHYGGVIK"
121589. .122029
/gene="ydcg"
CDS
121589. .122029
/gene="ydcg"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12283.1"
/db_xref="GI:2632776"
/db_xref="SPTREMBL:P96624"
/translacion="MNTNWIQVSEQHVLGKGAAGGFAQLCHGKKAPLAKMKEGDWLI
XSPRDADYDGLKRSFTAIGKSGNIYPYQMAPNFIPIYRLDIDYPCCHKIGFYDIK
SKLEFVQETKHLGFLFRGRHFEISKDFLTIAQAMGVNISGMAL"
join(122013. .122456,122464. .122479)
/gene="ydcH"
CDS
122013. .122456
/gene="ydcH"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12284.1"
/db_xref="GI:2632777"
/db_xref="SPTREMBL:P96625"
/translacion="MEWHSKFSGPNTSPCFLLWOATQSMORKVKGALAEFDLTHVQFV
LITCKYMIAGETVTOKKLASFSQTNIMVSEVVRTLEKGFIERSKNPQDKREVLL
SLTEIGGKVTAALPIVEKIDQAFPAAMKKENFSLGLQELLIKHE"
122464. .122479
terminator
/gene="ydcH"
gene
122550. .124904
/gene="ydcI"
CDS
122550. .124883
/gene="ydcI"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12285.1"

```

```

/db_xref="GI:2632778"
/db_xref="SPTREMBL:O31489"
/translacion="METSALLKQOIAKEIGLSQKHVESVIRLLEDGNTVPFIARYRKE
QTSQMEVQIQTISERWOYIQNLNORKEEVIRLIAEODKLITDNLKRIEQSVKIQEVE
DLRYPYQKBKTRATVAKSKGLEPLADYILTLQDOHLAATADOYISEEKEVFTREEA
IEGAKHIIAEQISDEPFRKWIQETFKRGTIKSAAGKSADTDEKNYMEYIEEPI
KVPVHRVLAWNGEKEDILKVAEPPADHIKAYLEKQIINKRSTSVREILQETIEDS
YKRLIOPAIEREKELSERKADQAIHIFSENRLKLLLPMPKGTVLGVDPAFRTGC
KLAVSDETGPKLVIIYYPHAPVNTKEAHEKVKKILQYQVEMVAIGNGTASRETEQ
FIVNLRDMPRKIYYIVTNEAGASVYASELAREEPPDLAVERSAVSIARRLODPLA
ELVKIDPKSVGCGYQHDVDSQKRLNLSLRFVETVNVQGVNNTASAALLQYVAGUS
KSVAGNVYKAREIGKFSNRKELKDIPRLGAKTYEQCIGFLRQEGTEPDLDRGTIHPE
SYKETKALLKLLGLSTEHIGTAELKDKINQLALSETAKELGIGEITLKDICEQLTRPE
RDRPRVPRPLKTLQLEDLKEGMELOGTVRNVYDFGAFVDIGVKQDGLVHSKLS
NQPVKHPLDVWSVGDIIVTVWVDGVQVQKGRVSLSMWK"
124887. .124904
/gene="ydcI"
terminator
125110. .125562
/gene="ydcK"
gene
125110. .125562
/gene="ydcK"
CDS
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12286.1"
/db_xref="GI:2632779"
/db_xref="SPTREMBL:P96628"
/translacion="MDNKLQKLTEDISETYFKKPFHQALFNDRLKTGCGRYLLTSH
NIELKNKYYIIEHREELIGIIKHELCHYHLHLEGKYGKXHRDRFRLMLQQOVNAPREF
PLKKAENKYYIYICTTCGQYIKKRAWNPDRYRCGRGKRIKRFIS"
125685. .125759
/gene="trns-Asn"
trNA
/product="transfer RNA-Asn"
gene
125685. .125759
/gene="trns-Asn"
CDS
125764. .125854
/gene="trns-Ser"
trNA
/product="transfer RNA-Ser"
gene
125764. .125854
/gene="trns-Ser"
CDS
125884. .125955
/gene="trns-Glu"
trNA
/product="transfer RNA-Glu"
gene
125884. .125955
/gene="trns-Glu"
CDS
125967. .126041
/gene="trns-Gln"
trNA
/product="transfer RNA-Gln"
gene
125967. .126041
/gene="trns-Gln"
CDS
126068. .126143
/gene="trns-Lys"
trNA
/product="transfer RNA-Lys"
gene
126068. .126143
/gene="trns-Lys"
CDS
126155. .126236
/gene="trns-Leu1"
trNA
/product="transfer RNA-Leu1"
gene
126155. .126236
/gene="trns-Leu1"
CDS
126317. .126400
/gene="trns-Leu2"
trNA
/product="transfer RNA-Leu2"
gene
126317. .126400
/gene="trns-Leu2"
CDS
126409. .126432
/gene="trns-Leu2"
terminator
126409. .126432
/gene="trns-Leu2"
gene
complement(126486. .127592)

```

```
/gene="ydcL"
complement(126486..127592)
/gene="ydcL"
/function="unknown"
/note="similar to integrase"
/codon_start=1
/transl_table=11
/protein_id="CAB12287.1"
/db_xref="GI:2632780"
/db_xref="SPTREMBL:P96629"
/translation="MSVKKLENGQYQVDVSFGDPITGERIRTPPVASTRKEALELEBA
KLAREFOEARKRSVSFTLISIYLASCEIDSKPNYQYQYIINKHISDYFLKSD
IQITHREITDFRKLHMETGLSNKSVNIMTSLSKIFDTAVHEILKRNPCDNVKKRLP
LTRKKFWPEEFKFIISLPQDLTKFTYFAFLTGRCEMLALQWKDIDKILL
EIDVHKSWLDGQFVVYTPKTNIRRVSKLKLLEWRKEAQOELFELGIRHS
HDYIFQYKTPPSRKDIFSRIKYECKDSLTPRLHDFRSHVALLIHQGEDYIYIK
ERLGHGSVKTTIDYGHLYPNKQKEMADKLDDL"
complement(127605..128114)
/gene="ydcM"
/gene="ydcM"
/function="unknown"
/note="similar to immunity region protein in prophage"
/codon_start=1
/transl_table=11
/protein_id="CAB12288.1"
/db_xref="GI:2632781"
/db_xref="SPTREMBL:P96630"
/translation="MTIYTSKGIKHKVQSVIKTHGTNNVYEICDIOKIIYILKNDLQO
ANGLLQHKATQQLIHINENLHQHQVIAHELGHFLHRLKNTFRVNVSKVLKDKL
EHQASIFASLLIUTKMLNEALPYIQGFSKEQTAAYNPVSFVTDYKLSIQGSFSNRI
YSHEISAFG"
complement(128111..128494)
/gene="ydcN"
complement(128111..128494)
/gene="ydcN"
/function="unknown"
/note="similar to transcriptional regulator
(phage-related) (xre family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12289.1"
/db_xref="GI:2632782"
/db_xref="SPTREMBL:P96631"
/translation="MSLGKRLKEAROKAGYTOKEAAEKLIGNNLSNYERDYRDPDT
DTLLKLSNLNVSTDYLLGKDEVSKNETDLNKTINEAIQELKDEDTLLFMNDGEFD
EETARLVKKALKNGIKFIDELKKE"
128768..128962
/gene="sacV"
128768..128962
/gene="sacV"
/function="regulation of the levansucrase gene (sacB)"
/note="alternate gene name: ydzc"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="CAB12290.1"
/db_xref="GI:2632783"
/db_xref="SPTREMBL:O31490"
/translation="MKGEFLTARDIQKILGVKOAKSYDIIRTNLNAQMKEGYVVIQOK
VSRAKFEECYCYGPKSQTG"
129273..129601
/gene="ydcO"
129273..129533
/gene="ydcO"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12291.1"
/db_xref="GI:2632784"
/db_xref="SPTREMBL:P96632"
/translation="MKIRVNATIIITDVEALKADKDLNHFKYATCFVVDADNHLKA
NEEISELLKGWDWMCINSTRPLSKDNQVFTLYNDDLYNLCR"
```

```
129572..129601
/gene="ydcO"
129503..130283
/gene="ydcP"
129903..130283
/gene="ydcP"
/function="unknown"
/note="similar to transposon protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12292.1"
/db_xref="GI:2632785"
/db_xref="SPTREMBL:P96633"
/translation="MNLNITVPDINMTFGDMKFMGLNRERYVDRENNKRTDVLESRI
YNIASVQGGQIEVTIPEYAGAKEIPPFADIELKNPKISAMATSQDSTYANVMWKLE
ADDIWVAGSGSVKPAATGGNEKK"
130319..131761
/gene="ydcQ"
130319..131761
/gene="ydcQ"
/function="unknown"
/note="similar to transposon protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12293.1"
/db_xref="GI:2632786"
/db_xref="SPTREMBL:P96634"
/translation="MSDFLNKRFWKYRKIRIPYRNKNVLAGALIEVPVLLSMFLF
WRELIHFDLSQVINKFENWVPLIKSVLCSVLIAVGSIVASYELLFDSYKKLIHROK
IARMFISNKEYENKVRKIFISNETDSKERITYFPMYQYQVKNHHIYIRIAMDMSRF
QNRFDLGDLNGLFCDLVQKMEGFVCFKLLYDVKKNRISIDDAVAENGVLPLMK
HISHQFKLPHMLJAGGTGGGKTYFMLTIIKACVGLGADRVILDPKNADLADLEVLIP
KVVYKSGNGLIMCLRKSVGDMERMDEMKNSTGTENVAYLGLKPVFIFFDYVAF
MDLLDMKERNALSYMOLVMLGRQAGYFLVLAGRQDKVYLAGIRQOFSFVSLGI
MSDTGYCMFGDVEKAYVNNKRGYANVGTGSLVEFYSPVPGYDFMSSINAKLV
GVEGAQATAVASGSVDQATASGEGVSEANG"
131754..132812
/gene="ydcR"
131754..132812
/gene="ydcR"
/function="unknown"
/note="similar to transposon protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12294.1"
/db_xref="GI:2632787"
/db_xref="SPTREMBL:P96635"
/translation="MDELKQPPHANRGVYIVKKEAVESPLVSMVDYIRVSVFTHDV
DRIIEEVLHLSKOPMTEKQSGFYGYGTIELDYIKFYSAFPDDRNRGLIEMSGGCRQ
FESPLECRKRTWYDFQDCMQQGSFTRFDLAIIDKKTYFIPELLKKAQKGEICISRF
KSDPFGSFDLSGITGCTTIYFGSKKEAYLCFYEKNYEOAEKYNIPLELGNWNR
ELRLKNERAQAVIDALLKTKDLTLIAQILNNYVRFVDADENITREHWKTSLPWSDFI
GDVGRLLPYVKPKQDFYQKSRNWLNRNSCAPTKMKVLEADEHLGKTDLSDMTAEALAD
KKMKMLDVYMAVDVDMV"
133077..133346
/gene="ydcS"
133077..133346
/gene="ydcS"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12295.1"
/db_xref="GI:2632788"
/db_xref="SPTREMBL:P96636"
/translation="MHNVNLLNQAGLENALASVGCLOIVEDLIEKMOEYVLYHTETA
EFAIDIFTVNNYTKVHAFNVLENEDEGETNVNKEFVHMHTE"
133385..133651
/gene="ydcT"
133385..133651
/gene="ydcT"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
```

```

/codon_start=1
/transl_table=11
/protein_id="CAB12296.1"
/db_xref="GI:2632789"
/db_xref="SPTREMBL:P96637"
/translation="MFNINLLNQARLESDDLSCGLDIAABELIEKMMEDFYSDVAQR
IIVOTLAVINNANAKLQATFNILDEDEMETDERSVEGESIYYE"
133668..133976
/gene="yddA"
133668..133976
/gene="yddA"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12297.1"
/db_xref="GI:2632790"
/db_xref="SPTREMBL:P96638"
/translation="MSAVDKRFKSLGFNVGIPSLVFVRSRDCMLVVEGERVKGFAE
YRYTFYTRVLPDGRMTSLKVMYENQSIKRVLHRVASFLSFLERTKQIEQKECKVAQ
"
133966..135030
/gene="yddB"
133966..135030
/gene="yddB"
/function="unknown"
/notes="similar to transposon protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12298.1"
/db_xref="GI:2632791"
/db_xref="SPTREMBL:O31491"
/translation="MLNKEIGSDKKESFFRSAPQKLKRIERPEKQKQKVPDRORSKLI
AVTLCVSGSLFICLLIARVQSLNRSQIANDMKDETNPNTNDDKQKISVTAAEANFLSG
FINEYMNVKRQMSLESYMYVQDNHDFEDERENVDGLGDRELKGYSLYN
VKEGDKSLFOYKYVTYENLYPVEKEVEKVDGKKKKKVKVKYKTESYEQMLNIP
VTNKGDSFVAVPYFTQIYDLKGDIAFKKEETRDEYAGEKKESIESFLQNFEEKYA
SEKKEENVYMKKPEALEGNLLFGEVQSVKIFETKKGFVFCVAFVRKEKENDIPVNEK
FSLEITENSQGFYVNLKHQ"
135042..135290
/gene="yddC"
135042..135290
/gene="yddC"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12299.1"
/db_xref="GI:2632792"
/db_xref="SPTREMBL:P96640"
/translation="MDPMNEFILGADLPTLGGVKGWASDVVVIQFITIVMFIAAKNLM
KLMGGIIFVCCIGSAVTWVIKHWSEFGWINALMEKL"
135303..135827
/gene="yddD"
135303..135827
/gene="yddD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12300.1"
/db_xref="GI:2632793"
/db_xref="SPTREMBL:P96641"
/translation="MGRLVENYRKAMREPKKIQOLPENYSLPFAVELIIPAINYFIFVG
LCGFWYGVRMIFPHAFDNSYIVIFGFFLTMVTKIRPEKNNIYIYFFDAKYTF
FIKLPRKKYCNDRKIDLSNEKQIEFRKLKLVVDYSNETKNAYEGNTQFEAVNKNRRY
GVLNKKQFDSYAK"
135715..138210
/gene="yddE"
135715..138210
/gene="yddE"
/function="unknown"
/notes="similar to transposon protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12301.1"

```

```

/db_xref="GI:2632794"
/db_xref="SPTREMBL:P96642"
/translation="MRLKTPMKAIHKNLLLRFTGVDWYAYYRIKSNIPMQNKEKVESEY
KKQWHLFEITSYEDFHLMMYPSEYELEKRFKOLETDAADAMDVARYNEETVRLLL
BORLGRLLTKYDFILGVLKLSLVNISVELKNILSFENTATDTVVVKMLGMEQNVSTSF
FEKYEEVEETLANIMASVGRGELSETEMYINRYHFVRGLKHQTNNESEIKDVRISITN
TIIDTPDPSVLHLSDQDEGYSAFVVIDEFLHNMSSEDLFYEAQSLPFPVEVQMKIQT
ESKITPALNKKRQOLKEEQEQSTGDRSDVSTVTSATMRIHLQDEIKKEDVHVNM
WLSVIVVHGKTKKECVKATIVKRHLKGAGITCRLPVADQLNLPFKMLPGGKLDITDK
NWIQTQDQVAESLFAVNSDIGKFGFLGWDRFOEHTDLESAIMSRDFVLPHFP
LANQLKSGSTRSPHCLITGDTGNKSYLAKLIFYNISMLNATKSLYIDPKKEMRKWIO
RVLNDVIRENFPYIAHLEKXNYITLDHENTHNGALDPSIFLPPMAKELVQVIFE
QVYDKGKDDIITAFKATSEVIDAKQGEQGVSLDIIRKMOHPPEAKVQAKAGDYLNE
VYDSILKLCIHGDSNPALSLEKITILEVENMDLPDHAERLENYISQLKSSAYMFA
LGKFCELFGMQNDQEQTFEIDEAMITFTSQQKVKVQRQMRIRGRSYNNAEYFISQSTK
DALKEEDSGNFGVAFADPEPNEREVELKMMNMEVTKDNKKMWESMFQGCCLFKDYVGR
TSKTSICLREEMOGALKTVKEKAVAAEYKL"
138229..138555
/gene="yddF"
138229..138555
/gene="yddF"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12302.1"
/db_xref="GI:2632795"
/db_xref="SPTREMBL:P96643"
/translation="MEIAFLNSLVVTSPTGFGYKAEKITLDEVKQWLKHVDGRYKSFIGH
KSTAQFLQKLLGIRIEQNRKTRFHMKYQKACFSLYERYPENVLLTQRDLEKARYQFY
LLTRKLD"
138559..141006
/gene="yddG"
138559..141006
/gene="yddG"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12303.1"
/db_xref="GI:2632796"
/db_xref="SPTREMBL:P96644"
/translation="MVYMKKRLIIVSAIVLLFTVASAVTVFSADGDTTTPQKVEKA
GGVELKVRFPISRYQANNEASDDLIRKAFVGLTNVTFSFAGNVRVVYDGTMDLIYNL
QPIDEFANSITNVSKTYIKTLKKNFGEALFIFTCAIVVYLFCVGRGSKVKAEMRRSILFI
CMVYIGGLMWSNAGYKMKVLNALSVEAGKLLTAGNLGVIGVDEGFAADSIAEKKG
EMEGTVAMRNLYFDIALMKPFLIVNDFETSEKKINEEDTDKGLNRIDKLLSYKLSE
DGEKDKDYIKETEIDDDYKNESMTSGNVNQLGESFTAVVASIVIGIPFLALAFNFI
LOVVALVIFVEPFAFLAVVPOLAYSQFVTLGRFLSGVILLKMLGVILVYVYTCFI
VDKLIIPNGFCMYLLNVANVASILMIGFHKRDAIIFKFTAGKVSVSDNNMMENMRQNI
VQPAWEQAKKIGVGWGGGVFTDFTKHFGRKDGSDNADVTGAPSGGSGSPSGTAMG
YDNTHAISRTPOKETANGIANHNSRSLKRNPTLSKEQEKOKOKEAFANAKENQQSH
LARLRKDGINSPLMKDALNEGEDLSKRAPILQDKDESARTDQKEVVEQILLQPNQO
QOTDDASLOHEESTSNRAPVLOENKEDTERTDQKAVIYDEONONLETDOQDDFEVOK
DDSVNSSEPVAOEKTAIEKSDOKVMNNOPEPOLGFESPOSTKYENQPLANNERKIRP
SEPAKVHSDGIRVDEKQAVAPAEKNKTVSREKQPSQTIKRTQSVNSFDQVSLNEIAR
RSSSKVEDRLRRDERTR"
141003..141992
/gene="yddH"
141003..141992
/gene="yddH"
/function="unknown"
/notes="similar to transposon protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12304.1"
/db_xref="GI:2632797"
/db_xref="SPTREMBL:P96645"
/translation="MISKVVLLPLVFSAPPIFFVLCIVVMTISRBNQVDDPIGGG
DGEYTVGIAPEVFRFAVEKARYOEGVFQDQVNIIMALTMOEGSGSLDIQMSSESI
GLPNSITDPERSTIEVGIKHFKKVFKQAGDVRLLTQAYNFGSGFIDYVKKNGKYYTK
KLALDFSRLOAFKMGWKSYGDPSPVDHVHRYKSGDKNKPVKGSMDFYETVWKALK
VEGQPYAWGSGSNPETGTCDSGLVQWSPAKAGITLPTAQEQHGATKKTISEKATAGDL
YFGGTYEGKAITHVGIYVGVNGRMFNNSDNGSIQYSDLKSGYWRDHLVVSFRIRK"
142007..142513

```

gene

/gene="yddi"
142007. .142513
/gene="yddi"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12305.1"
/db_xref="GI:2632798"
/db_xref="SPTREMBL:P96646"
/translation="MKTHIAWASACILLVMTGFTIGQOTYKLEIKDKNEVLSEKI
KELNHIESTSASENKAFFAFNYSDIDIRYETVKHKTGKGDDYAFPSRSDQKHIV
SVQSLSLSESYKAPLDESHELFUNIVEVATTANSVTITNQVLLIQTMTKKKEKDGWLVD
NQVKGNG"
142576. .142956
/gene="yddj"
142576. .142956
/gene="yddj"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12306.1"
/db_xref="GI:2632799"
/db_xref="SPTREMBL:P96647"
/translation="MKNLFIrLSLMMFVLTAACGSGSYDDAIDDVISQYKEHKNDTE
INIKRENAIVRVYEGGYIQFAFYPDNSRRELTTFKYIEKFGDKYEKMTDMPONGEN
DRGLSKTTPDYEEVKGETELEE"
complement(143147. .143947)
/gene="yddk"
complement(143147. .143947)
/gene="yddk"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12307.1"
/db_xref="GI:2632800"
/db_xref="SPTREMBL:P96648"
/translation="MDTLKKLFKFIANGIIGVLLDYSRPVLNTNIDQFRTSLENSDIE
LIRYSLQKLTWARNRENIYKNDEFNEFEHETENKIKKFLDLPVVDTEKSSIT
HFSSDONRELEKKIFISHSKDKIVCNAFVELLEDIGVSSRDIYTSPPYHGIGDED
IFEYLLKHLKFGAVFYMLSDNYDSVCLNMGATWVNSNCSFTFILPGFKGEIKGV
IDNKKAFSLSEEPIDLFNLKEKILRMYDLFLEDKWERIKAKFNTKLK"
144287. .145462
/gene="rapi"
144287. .145462
/gene="rapi"
/note="alternate gene name: yddL"
/codon_start=1
/transl_table=11
/product="response regulator aspartate phosphatase"
/protein_id="CAB12308.1"
/db_xref="GI:2632801"
/db_xref="SPTREMBL:P96649"
/translation="MRGVFLDKKIPYDLVYTKLNEWYTSIKNQDVFQAEILIKTEVEK
ELUNMEHQALLYQLLEFRHEIMLSYMKSEIEDLNNAETIKEIKEQGLTGMLE
YIFYFFGKFEFRKELISAIYRIAESKLSVEDEIERAEFPFYVYVYVYMKQTY
FSNNANRAKIFREYEEYAVQVRCOFIVAGNLIDSLYEERALEQFLKSLSEIKSN
IEHLIASHMNIGICYDELKEYKASOHLILALEIFEKSKHSLTKTLTTLTYVEAKO
QNVNALIYRKGFPIADKSDKEYSAKFILEGLFSDGETQLIKNAFYLASRKMF
ADVENFIEVADYFHEOGNLMLSNEYRYMRSEARKKIKGEIIDENQPDPSIGSDDFK"
join(145419. .145538,145566. .145591)
/gene="phrI"
145419. .145538
/gene="phrI"
/codon_start=1
/transl_table=11
/product="regulator of the activity of phosphatase Rapi"
/protein_id="CAB12309.1"
/db_xref="GI:2632802"
/db_xref="SPTREMBL:031492"
/translation="MKISRILLAAVILSSVFSITYLOSDHNTEIKVAADRGA"
145566. .145591
/gene="phrI"
145691. .146632

/gene="yddm"
145691. .146632
/gene="yddm"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12310.1"
/db_xref="GI:2632803"
/db_xref="SPTREMBL:P96650"
/translation="MGENMFKKXVTEYIWTILITITFIILSWGSYVNGTSTVSIG
QPTKVSQYITPINISPHDIKELRITFPQKLDVKQISSNEPINVKSDDKNIGVESNS
TFEIAKIVENNSVOLLITQKLNDRKNGNINSVYESQIVNPAKKOLINLI1
TSSYIFMLNIALIMKRWKDYAKMKEIKFEFEDNAKOLDKKSKSELSERKT
LNOAFETEDRIKYHEKKQOILLAKLNDYKKELTFWRNTIRKVLVELPDGDKKADKLI
GTVTSSLKTYGTEKNEHDYESLKVAALLNDSDKRS"
complement(147099. .148222)
/gene="yddn"
complement(147099. .147142)
/gene="yddn"
complement(147203. .148222)
/gene="yddn"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12311.1"
/db_xref="GI:2632804"
/db_xref="SPTREMBL:P96651"
/translation="MKKFDISVLSVAPRQGETMKQKIDSAVSLAKAVDNMGYKRIWF
AEHNHDAYASAATVSIVQHILANTKDIRVSGGIMLPNHSPILIVAEQFGLETLYPN
RVDLALRAPGTQDQADVIRRNHNGVFFFEREVNDILRFVGGKSVQGEVRAVPGIG
THVPFVLGSTDGSAETAALKGLPYAFGAQFSPHSMEEALSTYRENFQSPSYLOBPYV
IACINVIAAESIDEASPTISASHLOVYIDITYNNLSKLIPPTENFLESLSQFELETLHS
PLGYTMGDRETIIRRELIDFQOMYHADELIVLSNIELSKELQSYEILKQVVDLELKF
RWEQL"
148482. .148892
/gene="lrpA"
148482. .148892
/gene="lrpA"
/function="negative regulation of glyA transcription and
kinB-dependent sporulation"
/note="alternate gene name: yddp"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (Lrp/AsnC family)"
/protein_id="CAB12312.1"
/db_xref="GI:2632805"
/db_xref="SPTREMBL:P96652"
/translation="MIDEIDKILDELKSNRSLTMKKLGEKVHLTAPATASRVVKLID
NGIITKGSIEVNOVKLGFISHAFLNIYIEKITHOPYLAFIETQDNYVINNYKVSGDGC
YLLECKPSPNEVLDOFLNDLANKHANYKSVIVGR"
complement(149015. .149464)
/gene="lrpB"
complement(149015. .149464)
/gene="lrpB"
/function="negative regulation of glyA transcription and
kinB-dependent sporulation"
/note="alternate gene name: yddp"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (Lrp/AsnC family)"
/protein_id="CAB12313.1"
/db_xref="GI:2632806"
/db_xref="SPTREMBL:P96653"
/translation="MQIDSIDFOILQLLNKNARIQWKEIGEKIHMTOGAVGNRIKKME
DNGIIRKAYSIVVDELKMGFSFTAFVFFFMNAYMHDDLKLFATIRNEISEAHRVSGDAC
YLLKVTYHVSQEVNLHLLNDLLKGYQLYLSIKEVKHYNTLSMSDE"
149579. .150121
/gene="yddo"
149579. .150121
/gene="yddo"
/function="unknown"
/note="similar to hypothetical proteins"

/gene="yddm"
145691. .146632
/gene="yddm"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12310.1"
/db_xref="GI:2632803"
/db_xref="SPTREMBL:P96650"
/translation="MGENMFKKXVTEYIWTILITITFIILSWGSYVNGTSTVSIG
QPTKVSQYITPINISPHDIKELRITFPQKLDVKQISSNEPINVKSDDKNIGVESNS
TFEIAKIVENNSVOLLITQKLNDRKNGNINSVYESQIVNPAKKOLINLI1
TSSYIFMLNIALIMKRWKDYAKMKEIKFEFEDNAKOLDKKSKSELSERKT
LNOAFETEDRIKYHEKKQOILLAKLNDYKKELTFWRNTIRKVLVELPDGDKKADKLI
GTVTSSLKTYGTEKNEHDYESLKVAALLNDSDKRS"
complement(147099. .148222)
/gene="yddn"
complement(147099. .147142)
/gene="yddn"
complement(147203. .148222)
/gene="yddn"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12311.1"
/db_xref="GI:2632804"
/db_xref="SPTREMBL:P96651"
/translation="MKKFDISVLSVAPRQGETMKQKIDSAVSLAKAVDNMGYKRIWF
AEHNHDAYASAATVSIVQHILANTKDIRVSGGIMLPNHSPILIVAEQFGLETLYPN
RVDLALRAPGTQDQADVIRRNHNGVFFFEREVNDILRFVGGKSVQGEVRAVPGIG
THVPFVLGSTDGSAETAALKGLPYAFGAQFSPHSMEEALSTYRENFQSPSYLOBPYV
IACINVIAAESIDEASPTISASHLOVYIDITYNNLSKLIPPTENFLESLSQFELETLHS
PLGYTMGDRETIIRRELIDFQOMYHADELIVLSNIELSKELQSYEILKQVVDLELKF
RWEQL"
148482. .148892
/gene="lrpA"
148482. .148892
/gene="lrpA"
/function="negative regulation of glyA transcription and
kinB-dependent sporulation"
/note="alternate gene name: yddp"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (Lrp/AsnC family)"
/protein_id="CAB12312.1"
/db_xref="GI:2632805"
/db_xref="SPTREMBL:P96652"
/translation="MIDEIDKILDELKSNRSLTMKKLGEKVHLTAPATASRVVKLID
NGIITKGSIEVNOVKLGFISHAFLNIYIEKITHOPYLAFIETQDNYVINNYKVSGDGC
YLLECKPSPNEVLDOFLNDLANKHANYKSVIVGR"
complement(149015. .149464)
/gene="lrpB"
complement(149015. .149464)
/gene="lrpB"
/function="negative regulation of glyA transcription and
kinB-dependent sporulation"
/note="alternate gene name: yddp"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (Lrp/AsnC family)"
/protein_id="CAB12313.1"
/db_xref="GI:2632806"
/db_xref="SPTREMBL:P96653"
/translation="MQIDSIDFOILQLLNKNARIQWKEIGEKIHMTOGAVGNRIKKME
DNGIIRKAYSIVVDELKMGFSFTAFVFFFMNAYMHDDLKLFATIRNEISEAHRVSGDAC
YLLKVTYHVSQEVNLHLLNDLLKGYQLYLSIKEVKHYNTLSMSDE"
149579. .150121
/gene="yddo"
149579. .150121
/gene="yddo"
/function="unknown"
/note="similar to hypothetical proteins"

```
/codon_start=1
/transl_table=11
/protein_id="CAB12314.1"
/db_xref="GI:2632807"
/db_xref="SPTREMBL:P96654"
/translation="MTTALLVIDIQNDYFPNGKVALTNPEKAQNAAKLLSHFRNTGA
PVFVHOHTEGNIAHFPHFNPTEGVEIHESVRPLEKEFVIIVKHPNSFFNTDLNKGKQE
EGVKELVCGMSHMCIDATVRSAREVHGVCQVVEDACATTTLOIEDKIVPAEHVHYA
FMAALNGVIYATVTKTEAPLK"
150674..151438
/gene="yddr"
/feature="yddr"
150674..151438
/gene="yddr"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12315.1"
/db_xref="GI:2632808"
/db_xref="SPTREMBL:P96655"
/translation="MNTIQIRNATIVVYKANKFLIDPMLAEKGTYATFPETIRQDLF
NPLVSLPTSIDNILDGVDVAVIVTHLHLDHFDVAKNVLPKNIKMFVQNEADAKEVKAS
GFKNVEVLHGDVPEGIQLVKTGEGHGRGEBELNLMGDVGLVFKHPSEKVLVYAGDT
VWDAIEDETHHOPEIIVVNGGDNORLDLGLSLMGKEDIVEVHKAAPHAKIISVHME
AVNHTLISREELKNFSKEKGFSSNLLYPEDGESYTF"
151632..152942
/gene="yddS"
151632..152942
/gene="yddS"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12316.1"
/db_xref="GI:2632809"
/db_xref="SPTREMBL:P96656"
/translation="MVMTKKDTNISTOOPLWLOGYIDSPKONQLYKTKLILFISOI
FGAGGLGAGITVGAALLADMIGSENVAGIPTALTFGSAAVALLIGASSORFERRAGL
AGFLGGLGAIGVIVIAALNSVALLFVSLITYAGMASNLQVRYAGTDLANEKQRAT
AASMALVSTTLTGAVVGNLVNTMGFEADSIGVNLNAGPFIMSGAAAFIAGIILLIFLR
PDLFVSTAIANAEEKDDKVQIGGSKNPAIDKKGMVGAVIMILAQLIMTAIWMTP
VHGHGHGLSEVGLVIGLHIAAMLYPSPLTGLLVDFKGTMTAIASGATLLAAGLVA
AIAPADSLILLALVLLGVGNFGLLTGTALIIDSPHSLRAKTQGTDFVLLALSGA
AGGALSGMVVAHSSYTTILSIGAVLSLLLPVIVWIFRIQEKA"
153726..154449
/gene="yddr"
153726..154412
/gene="yddr"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12317.1"
/db_xref="GI:2632810"
/db_xref="SPTREMBL:P96657"
/translation="MRKKRVITCVMAASLTGLSLLPAGYASAKEDSKTTPSYEELAH
YKMKSEKISSNGKLVEYVSGNETHKVOMNGNHNHVKVDGIEQKGLNFEYDENVAKR
TNYENNKLKSNETTOAARPKGYHYVGLTSGHTKAKNALSVMSLVGLVPGWGWS
KAATILFTWAKEQIPDAYTKYDLYEKGAMTDSWYQATVQVFFEDKAHKKMKMKPMTS
TPAKVDLPNS"
154421..154449
/gene="yddr"
155372..155965
/gene="ydeA"
155372..155965
/gene="ydeA"
/function="unknown"
/notes="similar to glyceraldehyde 3-phosphate dehydrogenase
(glycolysis)"
/codon_start=1
/transl_table=11
/protein_id="CAB12318.1"
/db_xref="GI:2632811"
/db_xref="SPTREMBL:P96658"
/translation="MKKALFLIIDQYADWEGVYLASALNQREDWSVHTVSLDPIVSSI
```

```
GGFKTSVDYIIIGLEPANFNLLVMIGGDSWNSNDKKLLHVEKTAFOKNIPIAAICGAVD
FLAKNGLNNHSHTGNFYVLMKDYKQYKPISSFVEKQAVROKNLVTANGTAPIEFTNL
ILEMIDFDTNPENIEKMMYNNRYGFYHFCDKYGNPFVD"
156228..156470
/gene="cspc"
156228..156428
/gene="cspc"
/codon_start=1
/transl_table=11
/product="cold-shock protein"
/protein_id="CAB12319.1"
/db_xref="GI:2632812"
/db_xref="SWISS-PROT:P39158"
/translation="MEQGTWKFNAEKGFGERENGDDVFVHFSAIQSDGFKSLDEG
QKVSFDEQARGAQAAVQKA"
156449..156470
/gene="cspc"
/complement(157115..157576)
/gene="ydeB"
/complement(157115..157576)
/gene="ydeB"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12320.1"
/db_xref="GI:2632813"
/db_xref="SPTREMBL:P96659"
/translation="MFQIGDNIVYPMHGAGIIEAEKEFELEEKQOQYVIVRMSISNMT
VMPTSKILSNIRPVTIDILAKHMHIFQGESDRLLLPWKORYKINIDKTKTGEIQE
GAEVYRDLMRMKKALNASEKMLDNAYEFLISELEVIKITEKQIKSFG"
158112..158380
/gene="ydzE"
/complement(158112..158134)
/gene="ydzE"
/complement(158144..158380)
/gene="ydzE"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12321.1"
/db_xref="GI:2632814"
/db_xref="SPTREMBL:O31493"
/translation="MYLGIYSTACAFLLWNHGLQLLNASSGGLFFFPQPLVGTLLGW
LLGEQIGGTWIGSFLILSGVLLVKEKEVKES"
158478..159353
/gene="ydeC"
/complement(158478..159353)
/gene="ydeC"
/function="unknown"
/notes="similar to transcriptional regulator (AraC/XylS
family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12322.1"
/db_xref="GI:2632815"
/db_xref="SPTREMBL:P96660"
/translation="MKNFVIDQNLKELTEHRTVLPVACVKTITNQNINGYIPLHWH
EIQFVLILKGIALFQINEEKIEVHEGGDLFINSYLMHAEKEGSDCTYICLVNSPHF
VLSQELYSYVQPYMTSNLPYLFLAGNQWAKNILDVAKKINOLIOQKTSLEYIDIT
MQLTLMKNLVNNGFQLEYDOSEMIKSHRMKQMLNIIHLHYEKITILEDKAGQLSR
SECCRYFKRMLNKTPLRYVMDYRIQKSLLLQLHPESNVTVSVYQVGFNSTSYFISKFO
QAMNMTPLSVYKKNNS"
159466..160529
/gene="ydeD"
159466..160425
/gene="ydeD"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12323.1"
```

```
/db_xref="GI:2632816"
/db_xref="SWISS-PROT:P96661"
/translation="MNTSNRMGLFVITGATCWGIGGTVAKKLFOEYDIDWNLVLT
RLTTAGMLLLMQLPFRKDRSQVIEVWKHASKAGQLLIFGLGLMLAVQYTYMASIOHGN
AAVATLLOVLSQVILVIITLLRKQTVLTQODLITVVLAVGCFFLLTNGMSQOLSVA
AAVWGVLSGFAAAFTLYAVGLLNKFDLSLVVGNAMIIGGFALGFIIHPFWQDLFORL
TAEAVILFVLFGTMIAFWFFIKSLSPKETSLLGSLPLSAVVTVVVWLKAPF
GQFQWGAICIIIGITLILALHKEPSMESEKSIKIRNHANDI"
160504. .160529
/genes="ydeD"
complement(160578. .161450)
/genes="ydeE"
complement(160578. .161450)
/genes="ydeE"
/genes="ydeE"
/function="unknown"
/notes="similar to transcriptional regulator (AraC/XylS
family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12324.1"
/db_xref="GI:2632817"
/db_xref="SPTREMBL:P96662"
/translation="MDLLKSMNRALKYIEENLTNDIDFOEAAKLAFCSYHFKRMFSE
LAGISLSEVIRRRITLAAFELKDSNVKVIDIAIKYINSPDSFARAFONLHGINPSK
ARFNHSLKAYPRMTFQUTIKGNEMNRYRIEKEAPRIVGIRKRVDIIFKGINPETAS
MWESLDEKAINKLKSLSNVPGLISASTNFSSEGRTEENGELDHVIGAAATTKRCPDNF
SRLEVPASTWAVFESIGPPDPTLQEWGRIYSEWFPSSNVEQIEGPEILWNEHKDVT5
PSFKSEIWIPIYTK"
161668. .163099
/genes="ydeF"
161668. .163065
/genes="ydeF"
/function="unknown"
/notes="similar to transcriptional regulator (GntR family)
/aminotransferase (MocR-like)"
/codon_start=1
/transl_table=11
/protein_id="CAB12325.1"
/db_xref="GI:2632818"
/db_xref="SPTREMBL:P96663"
/translation="WRIMPINSFNPMSWKPSIDKAEPYIKALAGOLEQDILNGVL
LPGLTPPORELADYLDLNVSTISKAFKVCLEKGLLSATVSGSTFVSYDALSNAYLLE
DTPKHLETMGAILPDHASYEYDLPQLKNMVEEDYEKWFYSYGRAGESLWQDAAVKL
IQRGFTFESVDHILFANGGONIAATLASLQPGDRIGVDHHTYPGLKTVASLLSVQI
VPIKSENDMSPESELEYACKNENIKGLXLIPOYHNPTTSFMSVENRMRVADIVKYNL
FVIEDASVHLKNLPLALASAPQOVIHIASLSLAPGLRAYAVAPROYKEPLSK
ALYNLNIIVSPLAELTARTIIVSNQFEVLIESHREQTIIRNQLNVRIYAGITCLGVET
GIFRWLLLPGRMSGAEFEELAARLGVQVYAAERFVVGNSCSPERAVRSVCAPKTLBEL
EQGLMLRLRLDDL"
163074. .163099
/genes="ydeF"
163175. .164467
/genes="ydeG"
163175. .164467
/genes="ydeG"
/function="unknown"
/notes="similar to metabolite transport protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12326.1"
/db_xref="GI:2632819"
/db_xref="SPTREMBL:P96664"
/translation="MWLDKDSVKATDQVATOSLQSISSPEKOSLYKRTLFVVVISQI
FGGAGLAGVTVYCALLAQOMLGTDAFAGLP5ALFILG5AGSALIVGRLSQRYGRTGL
SAGFMVLTGGAIGVIMAAIINSIFLLFISLIYAGTATNLQARYAGTDLANHQRAT
AVSITWFTFGAVAGPSLVNMGDFALSIGVPSLAGPFIILAAAAYMLAGVILFIMLR
PQPLVIARTIEAANEPEPGDKHLAATEHTENKKGIIIVGATVMVLTQIVNVAIWTMPV
HMRHHGDLGAVGLVIGFHHIGAMVLP5LTVGLVDRLGRITMAIISGTTLLLAGVIAA
FAPGESMILLVIALSLGLGWNFGILSGTALIVDSPTDTAKTQGTVDVLIALSCAA
GGALSGMIVAGSSYLALSILGILLSLLLPVIVVWSRGR"
164626. .165142
/genes="ydeH"
164626. .165072
```

```
/genes="ydeH"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12327.1"
/db_xref="GI:2632820"
/db_xref="SPTREMBL:P96665"
/translation="MNHKWLSDLKKAGLALGKGVTILLKRFSLVIVTLMSIVVVL
AAAKSPGDHVISFDEPIILMISIGIVVFLIIPPLVM5FGNLVVRISGVYQCFTVF
TFGLIPIGFLIPNGFLTILVSITAGTLVSTASVAVTLCIGKNKKDI"
165125. .165142
/genes="ydeH"
165309. .165902
/genes="ydeI"
165309. .165902
/genes="ydeI"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12328.1"
/db_xref="GI:2632821"
/db_xref="SPTREMBL:P96666"
/translation="MTNSRTPKVD5EFLSKAKKKEEPEKLTILLDCELTEDFKMMH
PCYTHNKNIVLHGREYCALLFHKGVLLQD7DGLIIOQTENVQARQIRFTNVQEI
NELENILKAYIIEAIEVERKAGLVKNVKNIELNIPEELQNKDEIPALKIAFEALTPG
RQRAYFLY5QAKQSKTRESRVK5YVQKILDKGKLD"
complement(166254. .166913)
/genes="ydeJ"
complement(166254. .166913)
/genes="ydeJ"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12329.1"
/db_xref="GI:2632822"
/db_xref="SPTREMBL:P96667"
/translation="MKRRKICYCNTALLMILLAGTCD5KDG5AQOQ5NQA5AVQD
EKHTEPEESTKIRKDEAPIT5E5E5ATKAANDT5SAE5EK5EDNVLAAY5SSKIEYA
RWLQIPGNQOIEDELNVRIHAAGEPINPDNT5ASY5ENV5TOLAG5RLV5G5VYHGN
GDGTIHVNVPLRWD5ADDIEK5GMREVT5E5IKNRKTVV5D7DDEKIKRLDIDMI
H"
complement(167335. .168198)
/genes="ydeK"
complement(167335. .168198)
/genes="ydeK"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12330.1"
/db_xref="GI:2632823"
/db_xref="SPTREMBL:P96668"
/translation="MNKTTNGWINGFIGVLIFSG5L5PATRLAV5D5DPL5FLVCRAI
AGVLAGLLLI5FRQHPAKRDLISL5LVAF5V5V5G5P5L5TALALQHV5T5SAH5IVF5GL
PLATAV5FV5G5R5PR5PV5IF5SAAG5LL5VAG5FALIQGG5SP5L5GDAY5MLAS5I5VC
GL5YAE5GAKL5SRRLGNQV5ISWAL5LSPLML5PL5FF5T5PDS5SAIG5VPALL5LAY5S
LFSMLIG5FV5W5YR5GLAQGGIA5V5QLQLQ5P5FF5GLL5ASVIL5LH5EK5GVAL5VAV5IAV5I
MCV5A5ARR5PAK"
168353. .169744
/genes="ydeL"
168353. .169744
/genes="ydeL"
/function="unknown"
/notes="similar to transcriptional regulator (GntR family)
/aminotransferase (MocR-like)"
/codon_start=1
/transl_table=11
/protein_id="CAB12331.1"
/db_xref="GI:2632824"
/db_xref="SPTREMBL:P96669"
/translation="MDITP5L5NLT5DIP5L5VQO5LY5RYF5KEN5MHR5GR5TQ5K5MK5L5PS5KRLL
ANQ5ISQ5T5V5ERAY5EQ5LA5EGY5IV5K5PR5SG5W5FAD5H5DS5FAD5YR5M5PT5TP5IQ5E5AE
NKQ5ID5FHV5GNV5D5SY5FP5E5SAM5R5KM5VN5LDQ5Y5G5HEL5Y5R5PH5VL5G5FEL5RTL5IA5E5LY
```

QSRGVHCGPEQIIGAGNPILLQILCOVEFNPISIGVEDPGYPRARKIFEANRMNIVP
IPVDEGICIQIKEQOPNLVYVTPSHOFTLGITMTINRRLOLLKWAENQSFIEDD
YDEFRYTGQVPSLOGLDQHNRIYMGTSKSLPSLRISYMLPSPFLKKGHEITS
LYKTVSCHSOLILAEFIKNGEWMKHNRMKRAIVLEAVQRELGEHVIRGE
NSGRIILLDYLPLPGEKELIEKAKKHGKIYIPVSLSYQHHPPTKTVSLGFAGVSESDI
REGIKLKAWKI
169938. .170363
/gene="ydeu"
169938. .170363
/gene="ydeu"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12332.1"
/db_xref="GI:2632825"
/db_xref="SPTREMBL:P96670"
/translation="MKLDEFTIGQVFKTKSLKSKVDIMRFAGEFDPQYMHVDEEKAS
KGFNGIAGSIQTALISFKLWIEEGFYDDIIAGTEMNMTFKIPVYPDELFTIVE
VLDPKPRNELGILTVLLSTYNQEKVFEGLSVLIKR"
complement(170416. .170988)
/gene="ydeu"
complement(170416. .170988)
/gene="ydeu"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12333.1"
/db_xref="GI:2632826"
/db_xref="SPTREMBL:P96671"
/translation="MTQVYIIHGIRASSTNHWFPWLKRLLAGVQADILNMPNPLQ
PRLIEDLTSLYQHTLHENTYLVASLGCIPAILRLEHLQRLKGLGIIILVSGFAKS
LPTQLMDEFTQGSFDHOKITIESAKHRAVTASDDQIVPVSFSDKLAQIIDALYEVQ
HGHELEDEGFTSLPIVYDVTLSYFSKSTR"
complement(171070. .171399)
/gene="ydzf"
complement(171070. .171399)
/gene="ydzf"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12334.1"
/db_xref="GI:2632827"
/db_xref="SPTREMBL:O31494"
/translation="MNSLCRSKQAPFETLSLIGGKWKMRILYELGCEKTMRYGELKR
AMPFIHMLSAQLKELQDGLHRSEVSHPLKVEISLSDRGRSLYPLIDEMCKWGM
AQGGPHM"
171654. .172662
/gene="ydeu"
171654. .172526
/gene="ydeu"
/function="unknown"
/note="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12335.1"
/db_xref="GI:2632828"
/db_xref="SPTREMBL:P96672"
/translation="MNTPKRHKFRKAKMTLQIMVLIIGGIIAAYGLETVLIPNSVSDG
GVTLSIGVQSLNPLGILIAVINIPFVWLGKQIGKSFALISIIIGIVSLAAGTSFF
HHTPAIEGDTLLITVVGIIILGFGMLRNGGALDIDMLAVILSRKLPFGTSDLI
LFLNLFVTFVGLQCALLSVIATYASKVHVVEGLSGSKTFQIITQPELMV
ETIRDOGRSATYKEAYGGSHEKFEITCVINRLEETKLKELIIDIDKATFATPVYDV
AEVKSNGFRNLNH"
172637. .172662
/gene="ydeu"
complement(172676. .173062)
/gene="ydep"
complement(172676. .173062)
/gene="ydep"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1

terminator

gene

CDS

gene

CDS

/transl_table=11
/protein_id="CAB12336.1"
/db_xref="GI:2632829"
/db_xref="SPTREMBL:O31495"
/translation="MRNKLIGIDYSSDDFECPVETTLDIIGSKWKGLLYLHLDGKK
RFNEFKLYPKITORTMLTLOLRELERGVIHREYVKQPPKVEYSLTEFGTLEPVIL
HMKDGEKYKDRIDKLEAKKAEDKI"
173173. .173766
/gene="ydeu"
173173. .173766
/gene="ydeu"
/function="unknown"
/note="similar to NAD(P)H oxidoreductase"
/codon_start=1
/transl_table=11
/protein_id="CAB12337.1"
/db_xref="GI:2632830"
/db_xref="SWISS-PROT:P96674"
/translation="MDHMKTLVLVHPHNISSRINKKWEAVLSEPDVTVDHLYEKVR
DQPIDVEFEQQLLAHDRIVFPPLVWYSSPLLKQMPDEFTFGWARGPGGNLKGK
EWTAMSIGSPESHYQAGYNLFSISLTKPQASHLVGMTYLPSPFAEYRANTISDQ
EIAESANRYVYHITNIELNPKVRLQRYLKQLESVDLT"
complement(173910. .175097)
/gene="ydeu"
complement(173910. .175097)
/gene="ydeu"
/function="unknown"
/note="similar to antibiotic resistance protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12338.1"
/db_xref="GI:2632831"
/db_xref="SPTREMBL:P96675"
/translation="MANGKQAGQRISPSLTLLATACGIIVANLYYAQPLVGLISKA
IGLSPSAGLIVTQIGYVAGLFLVPLGDIENKLIWVSLLSAAALTFTAFVKH
GTLFLAAAFVGLGSIAAOVLVPFASVLSADARGVGVNMVSGLLIGLMSRPISSL
VADHGNALFALSAVSVILAIVLKVPARKPTANTNTALIGSKWKLRTTPVLR
RAIYHAFVFGASLFWTVPVLLSGDPDFHFSQKATALYALVGTAGAVTAPIGGRAD
RGLRLATGIALGVVYVSVLLPLMIQSSSPVGTIVAAAILLMDMGVSNALVLSQRAI
FSLAPEIRSLNGLFMAIFLGLGASIGGWAYAGSGWSTALWIGTAFVPIALLYFI
TEK"
175301. .175897
/gene="ydeu"
175301. .175897
/gene="ydeu"
/function="unknown"
/note="similar to transcriptional regulator (Tetr/AcrR
family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12339.1"
/db_xref="GI:2632832"
/db_xref="SPTREMBL:P96676"
/translation="MOSKRGRPRDEGTHKAILSAAVDLLEKGFDAVTVDKIAERAKV
SKATYKWSNKAAYIMDSFLSTATDRPLVPDPTGSSVQDVIYTHATNARLELTREGTV
IKELIAGQDLDEKLAEEYRTRFPQRLQAKLLEKGIQKRELNRDLIEVSDLIY
PIFYRLITGDEVNDSYVRDLVNMNFKVQATSATESM"
176505. .176840
/gene="ydeu"
176505. .176840
/gene="ydeu"
/function="unknown"
/note="similar to transcriptional regulator (ArsR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12340.1"
/db_xref="GI:2632833"
/db_xref="SPTREMBL:P96677"
/translation="MTIDVAAMTRCLKTLSDOTRLIMRLFLBQECVQCVQVDMFEMS
QPAISQHLRKLKKNAGFVNRDRRQWRYSINGSCPFDTLQLILHIDQEDLELNIHK
OKKTOACCO"
176853. .178549
/gene="ydfa"

CDS

```

176853. .178160
/genes="ydfa"
/function="unknown"
/notes="similar to arsenical pump membrane protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12341.1"
/db_xref="GI:2632834"
/db_xref="SPTREMBL:P96678"
/translation="MTSMILAVFIETLLVLVIVQPKLSIGWSACGGAVLALIAGVV
NFHDVLTGIVNATLAFVALIILSLDNIQFPFENALHNAKAAGYGVRFVYVS
LGLVALFANDGAALILPIVLAWRALHFNELVFPFIISGFIADTISLPFVVS
LNVNTVADYFHTIDYASRMVVPYLFSLASIIVLFFRKSPKRYDTEVKVPV
EAKQNNFRISWYILGLLIGYFASFEFSPVSVAGSIAIFIILAAQKSPAVHTKK
VYKEAPWIVFESIGMVVYVGVNAGLTDVLSVIOAAADQGLFAGTIGMCFIAIL
SSIMNLTVIDALAIAGTDPHGMREALIIYANVIGSDLGPKIPIGSLATLLHLV
LSHKGNKSWGYFKTGILTIPTLLITLVGLYIWLIIHSCF"
178519. .178549
/genes="ydfa"
178658. .179490
/genes="ydfb"
178658. .179443
/genes="ydfb"
/function="unknown"
/notes="similar to antibiotic resistance protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12342.1"
/db_xref="GI:2632835"
/db_xref="SPTREMBL:P96679"
/translation="MSVLQELTKKKYSLSLTKTFDDDKYCTFVYSILDQTPGAVYADD
QTFPKSFFTGESGIYFIAGDGNDFDFTAGYVEQVKSSKRTTLFSSDTPWSVL
KPLKDDLQNMERRAASFQPKSFKTLQLPKGLVKLRIDEDLIISHSTAFNSAYEYEW
NSYQFASKGFGFAYLGHNVHVSSECTSLFGLHNRAEMDIYTLLEERGLGLATCVANRF
IATCMENGVPSWDCDNCNSSIALAALGFKTVEITYEITYYSG"
179469. .179490
/genes="ydfb"
complement(179500. .180420)
/genes="ydfc"
complement(179500. .180420)
/genes="ydfc"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12343.1"
/db_xref="GI:2632836"
/db_xref="SPTREMBL:P96680"
/translation="MRKKTQVQIMISHLSMIGLWASAFFGIRAGLEGYTPHEHALFRLL
IGSMALLLFAVLQMLPDLKDIPIAIFLGLFGAFYHILLNIGKTVSAGVASLLVT
TAPIFSAMLSRLFYKHEFGFTKWLGSMTLSLGLVLIATGAGDFTYSMSGILVILLAAF
SESIYVFOARVYKKGRIPIPVFTFWGTIPMLVPLGLGEMMNASISATLSIVYL
GLLPVPIFALAYTSFVGASEATLSLVTPALALIISWLWIGEIPITLLSLGGVVT
VGGVCFTYLVKNMKNKIIFTHFSAKRVKKL"
180553. .182001
/genes="ydfb"
180553. .182001
/genes="ydfb"
/function="unknown"
/notes="similar to transcriptional regulator (GntR family)
/ aminotransferase (MocR-like)"
/codon_start=1
/transl_table=11
/protein_id="CAB12344.1"
/db_xref="GI:2632837"
/db_xref="SPTREMBL:P96681"
/translation="MSFPEEKMKKLPKYRQIVHFIFKEKIGNGEWPIGSKIPSQRTLA
KQFQNRSVITALELMADGLIEGTMGKTVVINNTWTLMKNSAPDWDQVYTSQIO
MPSRKIVOPINQESNTDLOLSKGELSAEIPFLAVMKEMCKVSOHMEAFGEYEPKG
YPLREALSNLYKTGINVSSSSILVSGALQQLISMLGLQSGTVYLDQPSYLS
LHVFSAGMKLGLPMDNEGLLPENVHLTRGERRAILYINPCFNPTILMSKKRRE
ETLAVSENQLPIEDDIYRELWIDEIPPIKTIKNGHVLYIGLSKTLSPGLRIG
WIVGPEPVIERLSDIKMOTDYGSSSLSRVAAEWFTSGHYQQHVKEKVRQOLKVRRELA

```

terminator

gene

CDS

terminator

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```

LSAETHLKEVATWNPKGFFVWIKILPISIMKLYTKALSKGILINLSGSIYAQSKG
NYIRLSYAYASLEDLQKGIYELGLMIKELASR"
complement(182119. .182742)
/genes="ydfE"
complement(182119. .182742)
/genes="ydfE"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12345.1"
/db_xref="GI:2632838"
/db_xref="SPTREMBL:P96682"
/translation="MDIKKQDTRFRKEIKPKIMYGTSTFLLTLNEDGTTNISPMSS
SNALGHYILGVLGGKGAJNLEHRKECVINLPDQWENVERISSYSGKKSIPPLKK
QIGFYTKYKGEAAGTLPQSKTVPSTRICEPIQIEAEVKHRLRPEYESSFAIVETQ
ALHFAEESIILDENHINPWSPLYNFRHFGVGLGREGVKTFRSET"
182832. .183512
/genes="ydfF"
182832. .183512
/genes="ydfF"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12346.1"
/db_xref="GI:2632839"
/db_xref="SPTREMBL:P96683"
/translation="MNPNTAKISSLLSDPSRSSILLSLMDGRHPAGELAYLANIKPO
TASFHLNKLLEAKLISVEKHGRHRYRLADSDAANVIEQLLHIAKPAKYSKESKEK
KDLHFARTCYDLAGYGVQMAHSLVEQGLMKVLDNFVETSEGALFFSKLGLINESQF
RNRKRAFACCDLWSERQHIAAGALLVRLMEEWIVRMPKTRAIRITOSGKIATF
EKHLNVKI"
complement(183592. .184035)
/genes="ydfG"
complement(183592. .184035)
/genes="ydfG"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12347.1"
/db_xref="GI:2632840"
/db_xref="SPTREMBL:P96684"
/translation="MSQRYVYEIAPEGMKIMMDEKYTKQSSINRTTRELKIRVSO
MNGCAFCIDMHTSDARKMETEQRIYCLHAWNECDFYSPPEAKALSELHITLIPSKR
VPDELVHRVREHYDEQYVDLVLIINQINSWNRISTAMGNRAASK"
184708. .185931
/genes="ydfH"
184708. .185931
/genes="ydfH"
/function="unknown"
/notes="similar to two-component sensor histidine kinase
[ydfI]"
/codon_start=1
/transl_table=11
/protein_id="CAB12348.1"
/db_xref="GI:2632841"
/db_xref="SPTREMBL:P96685"
/translation="MLIRNPFKDYSHDRALNMLALVPGLAFITLMYIASIVLQF
VSGWSILULYATFIIAIFALLHWHYSRWKRVLYIFAVQGLITFALANLMTGFI
LVIIGLYAFILQIGMADRRTFLIYLILLVINSAYHLHKGVELHFVITAAPIMI
VIITTAATFAQVDEKIKQAQTLERLELAHQOQVQLTQNRQRMARDLHDTLAQGLV
SLNMOLDATHVLAKGNTERAKEIIOQSMKRVKSTADARSADDDRKSSEEIGVLKE
RITSLMDHFIESTGMACLLDYRLHOVLDRTAENCYIIGECMTNAAKHAETKIWS
IWNDEKRLHLVPKNGKGFDEKGGKKRGHVGLLGIQERVRAINQFNKSTKLKGT
QIEITVPIQEMQDE"
join(185924. .186565,186591. .186610)
/genes="ydfI"
185924. .186565
/genes="ydfI"
/function="unknown"
/notes="similar to two-component response regulator [ydfH]"
/codon_start=1
/transl_table=11

```

```

/protein_id="CAB12349.1"
/db_xref="GI:2632842"
/db_xref="SPTREMBL:P96686"
/translation="MNKVLVDHLLVREGKLLIETNDQYTIIGEAEKQVAVRLAD
ELEPDIILMDLMPMSGLEAIQIKKEKHDTPIIILTYNEDHLMIEGIELGAKGYLL
KDYTSSETFLPTMDAAIIRGNVLLQPDILKRLQEQIOFERMKKORNETQLTKEVIVLKAI
AKGLKSAIAFDLGVSBRTVKSLRTSYNKLANSRTEAVTIAMQKGLITDN"
terminator
186591..186610
/gene="ydfj"
186681..188855
/gene="ydfj"
186681..188855
/gene="ydfj"
/function="unknown"
/notes="similar to antibiotic transport-associated protein"
/codon_start=1
/transl_table=11
/protein_id="CAB12350.1"
/db_xref="GI:2632843"
/db_xref="SPTREMBL:P96687"
/translation="MSKMLYTLGQVARNRIKALCAWIVLVVAIGLAVTLKPSFSD
MSIPDPSEKAMDVIOKEFPHPGPKGSIRVIFGAGGEKLTGRPAKKAIEDTFKEISK
DDSVDSIASPFVTGTAKDGTAYADIOYKSSADDIKDYSIKHLKDSLKMADDEGLQT
ELSGDVPGAEMIEGGVSEIVGIIILAFVLAITFGSLLIAGLPILITLIGVSGIVL
IGTVQVEDIASVSLAGMIGLVIDYALFIFTKHQFVLGEGIQKNESIARAVTAGS
AVVFAGLTIVLALCGLTVNVPMSAMGLTAGLSLMAVLASITLVPVLSIAGKRM
PKSNKIEKQSTETNWRGVTKNPMLSVCSILILIVISIPSMHLELGLPDAGMKAK
DNPRRAYDLAEGGEGFNGQLTIVADATNATENKAEAFADAVKEIKGLDHVASVTP
AMPKRGFNLAITVPETGNDVTKDLVDHVSRLSKDNGVDLLVLTGSTAVNDIDSR
LNDAIPIVFAVLIIVFVGLITVFRSLVYAVAGFMLTMTATLGCIVPVLQGNLI
DFPKIPEKGPLAFPLILSIGILFLAMDYQVFLVSRMREYVKTNNPQVQIAGLKH
SGPVVTAAGLIMIFVFAETFEAGEASTKANGALSFGVLFDATFVMTLTPSVKLMG
NAWYIPLKWLDKLIPNVDIEGHQLTKEIOEIDHEQKKQISV"
/gene="nap"
complement(189267..190169)
/gene="nap"
complement(189267..190169)
/EC_number="3.1.1.1"
/codon_start=1
/transl_table=11
/product="carboxylesterase NA"
/protein_id="CAB12351.1"
/db_xref="GI:2632844"
/db_xref="SPTREMBL:P96688"
/translation="MSNHSSIPELSDNGIRYYQTYNESLSLWPKVCKSFYISTRFQG
THVIAAGPDAPLVLLHGLAFSPMSYPNIAADWSKYRTYAVDIIDGRNKSIPENVS
GRTDIYANMLDVDFNLGIEKSHMIGLSLGLLTMFLNMPERVKSAAILSPAETFL
PFHIDFYKALGLTASNGVETFLNMMDQNVLHPILFVKQKAGVMQDGSKNPQNA
DGFVYFTDEELRSARVPILLLLGHEVIYDPHSALHRASSFVPDLEAEVINKAGVL
SMEQPTVNERVMRFNAETGISR"
complement(190371..191060)
/gene="ydfk"
complement(190371..191060)
/gene="ydfk"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12352.1"
/db_xref="GI:2632845"
/db_xref="SPTREMBL:P96689"
/translation="WFGIFNVIIAGSIIGGIFKGIKDEYQDILMQAMGAFAVAL
GINAIOHLPDSKYPIPLFIVLSIAGLIGLQIINLELRNKLNVKFSKNLAELGSTAV
LLFCIGSLISILGPEAAALHGDYTLITNGMLDGTISIVLASTFGFGIAAAALVLSWQ
GSIYLFQVMYESAINTDLINETIIVGGIILSSGLSILGKFKTLNLLPSLLIPPV
IFVIAFGLRF"
complement(191150..191962)
/gene="ydfj"
complement(191150..191962)
/gene="ydfj"
/function="unknown"
/notes="similar to multidrug-efflux transporter regulator"
/codon_start=1

```

```

/transl_table=11
/protein_id="CAB12353.1"
/db_xref="GI:2632846"
/db_xref="SPTREMBL:P96690"
/translation="MKBFESIGEVSKLFNVKISALRYVDEIGLLKPEYKDEQTVRYR
STQOFERLDIKYLRALGLPINKLLDYPNSONTWTLHLKASQOVEDIRKKKLELERIE
RKISRRTHIEDAVNMPHLHQISKIKLPAMRVAYLOHEYVLGHDIHSLAEIRLTHLVN
EIFRIGKIGISISAANKAKQFDKYSIFMILEDENEKTSSEIIFPSREYLIQIRFGS
HPEAPYKKLLAYMKHHYEVAGDSIEITLIDYGITNNLDNVVTEILLPIK"
complement(192052..192966)
/gene="ydfw"
complement(192052..192068)
/gene="ydfw"
complement(192073..192966)
/gene="ydfw"
/function="unknown"
/notes="similar to cation efflux system"
/codon_start=1
/transl_table=11
/protein_id="CAB12354.1"
/db_xref="GI:2632847"
/db_xref="SPTREMBL:P96691"
/translation="MASERQISRKVKALIALIANLILMAGKVPFGLVCDSEAVFADGI
HSAADVASTAVLAVIGISNKPDPQDPHGFCHGAEVISEAIVGIIILVIVSYILIEAI
LSFVKGPSQYISALFAALISYVAKIELIYRSIRQKKKNSKAIILAIYDHKGDIVAS
LAAPIGVLLAIIIGNSRGSYLLYADAJASIVAYLIPKISMELIRPSGVLMKESVDP
ELIEEYKAVIFQCDQVKRIDIRAREHGHYKLLDVLRLSDHDLTKQGHDIAREIRNE
IKRQFSDEVEVLHVNPYEE"
193442..194062
/gene="ydfn"
193442..194062
/gene="ydfn"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12355.1"
/db_xref="GI:2632848"
/db_xref="SPTREMBL:P96692"
/translation="MAEFTLVNERSASNFLSGHPITTKEDLNMFPFLVALAPAFNL
OHTKYVTVLDQDVKEKIKQAANGQYKVVSSAVLLVLDKQAYQQARADIYEGCLKVLI
LNKQEYDHWQDVTVPYENRGEQFKRDEAIRNASLSAMFMLSAKKMGWDTCPMIGFD
AEAVKRLINDDQFEVMMITIGKTESRPRGRKPKVNEFVEYM"
194078..195016
/gene="ydfn"
194078..195016
/gene="ydfn"
/function="unknown"
/notes="similar to ABC transporter (binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB12356.1"
/db_xref="GI:2632849"
/db_xref="SPTREMBL:P96693"
/translation="MAKMTGMHHIHTAIVGHPQENTDFYAGVLRLVKQTVNFDPPG
TYHLFNGEGKPGTITTEFPWAGARQGVIGQGVTSVYVPGAMAFWEKRLKFN
VPYTKIEREGEOYVEFPDPLHLEIVEREGEANTWTEGVTDDVAIKGFGGATLLS
EOPDKTADLLENIMGLERYKSGEDFVRSAGDIGNVIDLKLTPIGRQMGAGTVHHI
AMRANDEDQLDWORYIASHGIGVTPVRDRNTFNALFYREHGEILFEIATDDPPGFAHD
ETQETMGEKMLPQVYEPHRTQIEOGLLPFEVRELD"
195118..195551
/gene="ydfp"
195118..195507
/gene="ydfp"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12357.1"
/db_xref="GI:2632850"
/db_xref="SPTREMBL:P96694"
/translation="WEDAGLLIRIMIGVVFLFYGTQKLFGWFGGYGKGTQGWFEESI
GVKPGNAAALSGELVSGILFVLPLGAAIITIIMLGAIVKVGAKGFANGANG

```

```
terminator
195523. .195551
/genes="ydfp"
195693. 196065
/genes="ydfq"
195693. .196031
/genes="ydfq"
/function="unknown"
/notes="similar to thioredoxin"
/codon_start=1
/transl_table=11
/protein_id="CAB12358.1"
/db_xref="GI:2632851"
/db_xref="SPTREMBL:P96695"
/translation="MKEMTGLSLVSTENFIKHKFSFIVISRPGCVCHAVLPOLRI
VLQDFPNIKIGHINADDDVAEAGRFSVFTVPVLLLFIDGTEFLREARFVHFQLEQL
KRYLYEGE"
terminator
196050. .196065
/genes="ydfq"
complement(196071. .196307)
/genes="ydzH"
complement(196071. .196307)
/genes="ydzH"
/function="unknown"
/notes="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12359.1"
/db_xref="GI:2632852"
/db_xref="SPTREMBL:O31496"
/translation="MRSQINSLNTINGMTQNANTDIGQLNSHTANSKNYGVNFTL
GDYSPSHSIIISASCDNDTSDQGVDNPSPSEE"
complement(197193. .197870)
/genes="ydfR"
complement(197193. .197870)
/genes="ydfR"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12360.1"
/db_xref="GI:2632853"
/db_xref="SPTREMBL:P96696"
/translation="MNRPEGYIKRLAGRKANPMTNLIHHWGGSMNFTWESLVLIIV
AGVLLIRISGRKSAQMTSTQTVVMISIGTIIYQPIEYSLEFKTLIAAIFTSLIVM
EWQMSKNTTEKMLTGKAKIVENGOLHIENLKKMRLTADQLEMOLRLHGVTAQDVK
IATLEANGOLGIELTDDAKPLTVRDLKKLHPDFINKDGAQSQSQNQNFDEVGKKNK
NPKKLH"
197983. .198690
/genes="ydfs"
197983. .198690
/genes="ydfs"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12361.1"
/db_xref="GI:2632854"
/db_xref="SPTREMBL:P96697"
/translation="MIELEVIVRTVASFGLLLIAERILKQOTISQMTIFDFIAITLG
AIAAGLAYNTSIKPHNMAISFSIFVLTIFLISIKNRKLRKFFAGDPTVLQNGKI
LESNMRKRYTLDYLNQOLREKEIFNEIEVFAILETNQGLTVLRKPQFRHVTQDILM
IAYNQEORLPIELIMDGEIIEENLKNQLRFESWLLELRKIDIKVKETVYAVLLGND
IYVDQVKDHTSVPMKE"
complement(198705. .199136)
/genes="ydfI"
complement(198705. .199136)
/genes="ydfI"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12362.1"
/db_xref="GI:2632855"
```

```
gene
CDS
/db_xref="SPTREMBL:P96698"
/translation="MDFEIRKWLIEITNEYKQSDFTNVLKYKAPHEHFFDSEASTFVY
DFYODEEYNFTIIVPGVYREKLTIRLLSKTQLLIKGTITPVPFAEMEVLREYIGEI
ERIIQLPEAAETHLLQIOLLNGLLHISYPROVEIVAFNKG"
complement(199149. .199391)
/genes="ydgA"
complement(199149. .199391)
/genes="ydgA"
/function="unknown"
/notes="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12363.1"
/db_xref="GI:2632856"
/db_xref="SPTREMBL:P96699"
/translation="MPYQINIANIKINGVTONGNIDVCPVHNSHTANSKYFGANFSL
GDLSPSSLNTGNIDSDVDQDIGNPSAFISNQI"
complement(199405. .199677)
/genes="ydgB"
complement(199405. .199677)
/genes="ydgB"
/function="unknown"
/notes="similar to hypothetical proteins from B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="CAB12364.1"
/db_xref="GI:2632857"
/db_xref="SPTREMBL:P96700"
/translation="MPSTVINLYLKINSNGSITIGEAAYSNPNTNKSQGTNSS
FGDTSPTSESVNENFLNDPDVNDQTSIGNSDTSNNAPIAPPILD"
199976. .200563
/genes="ydgC"
199976. .200563
/genes="ydgC"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12365.1"
/db_xref="GI:2632858"
/db_xref="SPTREMBL:P96701"
/translation="MKSQGESRRRLKAAANESVRFHDAKYSEIVKKAGTQPSF
XLYPQSEAFIAELITDFHSVRKLTSLLENGLTEDVSKRVLLAVETVQFLDED
KLTFKIGFFLNPEAKQKDLAMVKNLEAEQRLGYFHSLEDMETVAECLVGMIEHL
TESLLTGIDKDPASIAAEVNNLIYGMLPKGNDRV"
join(200560. .200904,201024. .201041)
/genes="ydgD"
200560. .200904
/genes="ydgD"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="CAB12366.1"
/db_xref="GI:2632859"
/db_xref="SPTREMBL:P96702"
/translation="MISIMKVSIAVFMLAGGIKVSVPFQVEHRRHYQYPLWFLTV
TGILEIAGALAMTAGIWNRYAAGVLFVVLMAIAHMFARQSVIMAQAMICL
IVSIIIMGSYT"
201024. .201041
/genes="ydgD"
201067. .201540
/genes="ydgE"
201067. .201540
/genes="ydgE"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB12367.1"
/db_xref="GI:2632860"
/db_xref="SPTREMBL:P96703"
/translation="MAFTLEDTEEFEEFAFRGMSVQNFPAKONITSGTWTEKAEKSE
QAYENMIPNGRDSNHYFWNITNDQGRMGWLWLYADPLHPQKEAFIYSFGLYEAFRG
```

```

KGLAQLALQTLDERARALGAERLALHVFAHNETAVLYLQKMGYAMTILMRKQLC"
complement(201655. .203343)
/gene="exp2"
complement(201655. .201672)
/gene="exp2"
complement(201700. .203343)
/gene="exp2"
/codon_start=1
/transl_table=11
/product="ATP-binding transport protein"
/protein_id="CAB12368.1"
/db_xref="GI:2632861"
/db_xref="SWISS-PROT:P39115"
/translation="MKELIVTLNVSVEYKQOTVFKHVNASVQOQDIIGIKNGAGKS
TLHLHINDAPAQQLRKDKIDKALVQEATAAYSFADQTPAEKKLEKWHVLRDPH
QLSGEKKARLAKSEDADLLLDDEPTNHLDEKSLQFLQQLKHQNGVILVSHDR
YFLDEAKTKWLEDDQTLIEFGKNYSYMKFRKKRLTOQREYKQOKMYERIEAQN
GLASWSKHAQSTKKEGFEYHRVKAKRTDAOIKSKQKLEKELEKAKAPVTPYVT
VRESIDTHKTGKFEVQNVTKAFGERTLFKNAFTIOHGEKVAIIGPNQSGKTTLL
NIILQETAGSVVWSPSANIGYLTQEVDFPLEQTPPEELFENETFKARGHVNLMRH
LGFTAAQWTEPIKHMSEGRVKIKLMAYILEEKDVLILDDEPTNHLDLPSREQLEETLS
OYSGTLVASHDRYFLEKTTNSKLVISNNGIERKQNDVPSENEREREELRLKLETEROE
VLGKLSFMTNDKGYKELDAQNELTKRIKELDHQKKD"
complement(203378. .205039)
/gene="ydgf"
complement(203378. .203400)
/gene="ydgf"
complement(203628. .203653)
/gene="ydgf"
complement(203663. .205039)
/gene="ydgf"
/function="unknown"
/notes="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB12369.1"
/db_xref="GI:2632862"
/db_xref="SWISS-PROT:P96704"
/translation="MTDDMTKDNINOOTLQRLKKNRHIQLIAGGAIGTGLFGSGKS
IHPAGSILFAYMITGICFLIMRSGELLNLNYSFVDFVDYVLDGMAAFITGWT
YFNCWISIAMDLTAVGLTYWLPQVWPVGLIALIILLIMNLATVKLFGLELFWF
ALYKTAIALLIIVIGLVMIFKFGSSVSFTNLASHGGLFPGNHGFTLSQMVVF
AFVGIELVGLTAGETENPEKVIIPKAINNIPRVLLFVIGALLVIMSYIHPWDINPES
PFVQVFVAVGIYGAASIIINFVLTSASACNSAVFSTRMVYSLAKDHNAPESMAKLT
QRVPARNKFFSAIIVILIGVTLNYIMPEGVFTLTSISTVCFIYIGVITVICHMYRK
TRPELAKTNKFLPLYPFTNYILAFVVLVLAQDTRVSLFVTPVPMFILLIYV
KVRKAKHQ"
complement(205185. .205729)
/gene="dinB"
complement(205185. .205206)
/gene="dinB"
complement(205211. .205729)
/gene="dinB"
/codon_start=1
/transl_table=11
/product="nuclease inhibitor"
/protein_id="CAB12370.1"
/db_xref="GI:2632863"
/db_xref="SWISS-PROT:Q02886"
/translation="MSDFAPKLYEVNWMANQOIFNRLKELPKPIYHQEIQSVFPPSISH
VLSHVYSLDLGTEVFPFGKTLSDALAEOLKEOTEAKEIEEMEDLFLLSERYILFL
QOKEOLNKLPIQIONPSSGIMTKTVSELLPHVNVNNGHYHGRNITAMLRQAGIASAPTDY
GILFPMKTEKA"
205898. .206356
/gene="ydgf"
205898. .206356
/gene="ydgf"
/function="unknown"
/notes="similar to transcriptional regulator (MarR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB12371.1"
/db_xref="GI:2632864"

```

```

/db_xref="SPTREMBL:P96705"
/translation="MNEQKLCQAINLFVEVLEGTFEVHREINQDVFXHISREQADLL
KILVKGPTSPGSAONVHKSAISNRKLLKLEKGLVQWDDCEPKSRRKRLNITP
SGEHILELDSAIFNALKPLDIDDEHLCHEIIFTLKSKFKGGSAAE"
join(206353. .209010,209020. .209041)
/gene="ydgH"
206353. .209010
/gene="ydgH"
/function="unknown"
/notes="similar to transporter"
/codon_start=1
/transl_table=11
/protein_id="CAB12372.1"
/db_xref="GI:2632865"
/db_xref="SPTREMBL:P96706"
/translation="MRAIIFKWAIAITVALTVVLSLSPNLTELANOKQAOLPAD
AVSERNALIKQAGEDNNSISVFTLONAIIKKTENOLRIIDKIKKIDGVEEVTSP
SAKEVVDQLMSKDKKTVLMPVITIGSDKKAETADEIYQIVPDDLTAITYTASLINQ
DFAPHSSEGLKKTETIVTCLIGILLIVFRSVVTPFPIVVVGFYSILISQILGILVY
NVDPISFTQTFLVAIILFGITDYCILLTRFREELANGHDKKEAALIAIYATGCKTL
FISGFVILIGFSALGFARFAIFQSAVGAVGVGILMIIYTLPLFMVTLGKELFWPS
KKVLSHSDNKLWAFGRHSVARPFLTVITVITLPLFILTVDDOIISPDSTAEISSDYK
SIKALEAIKDGEGEGKAPLINVVVKDKDLTADTIPYLGNIKAIKDVHDVSVMTI
TQPTGKKIKDLYIDNQLGSVSDGLDKTVKGIADVQSGSLTDIENGLNOMAGTGSASNG
GSGSLGDAADGLGKINQQLVLSKQISOTGNTAQTVQOQLTAISQQLGQIQITGLEQAN
QQLSGOQAAGTILTESLKLESEGVKSANEGTLKVSNDGITASSDILEDMSKSPVTRDTG
IFIPDQVMKDRFKSIOYSFADKGVQLSVLDNSPNYSQAQITTIQIKKAVANEY
DGTPLENAQIVYGGVTSMNADLKELTSTDFSRVIMTIGLFIYTLTILFRSMIPYIM
IASILLTYTTSISITELIFVNGLGNAGVSNVPPFSEVILLIAGLDYSIFILLDFKEE
VHLGIEQGVVRSKMGSVIITAAIILAGTFAAMPSPGVNTLMQVAVIIIGLILLYGL
VILPLFIPALIIATFEGENWMPFGRKKGKE"
209020. .209041
/terminator
/gene="ydgH"
BASE COUNT 62215 a 41697 c 47649 g 57539 t
ORIGIN
BSUB0003 Length: 209100 February 15, 2002 13:44 Type: N Check: 9816
Found using 'seq4' (pappu403.key)
...
43999 ACACGGCGCTGTGATCAATACAGCGATGCGGATGCCACTCCGCTTTCAACGTTCCGC
|-----|
44049 44056
44059 TGATTCTGTGCGCGCTGCTTTCTGCTAAGACGAAATGAATTGCGC
...
98926 CTTAGGCTCTTTTATGAGGAGATTGCTTTTTCGGAAGCAAATTTGCAAGACGTTCCGC
|-----|
98976 98983
98986 AAGGTGCTCGAGGCTCAATTTATCATTTGCGCTCGTCAACTGTATTCTG
...
156462 TAGGGCCTGTTTTTTTATGCTATATCATCTACTCAATAAAAAACCCCTACTCAACGTTCCGAG
|-----|
156512 156519
156522 TAGGGAGATGCTACAAGGTAAATAGAAAAAGATTTAAACACTTCAAAGGT
...
-----
1 match found in sequence:
cvgr190 ; TOIG of: cvgr190 check: 7467 from: 1 to: 52283
(from "immun_ge.seq")
TOIG of: cvgr190 check: 7467 from: 1 to: 52283
LOCUS CVGRI90 52283 bp DNA VRL 29-APR-1998
DEFINITION Cowpox virus strain GRI-90 DNA (52 kb fragment).
```


ACCESSION

VERSION

KEYWORDS

Y11842
Y11842.1 GI:3096962
C10L gene; C11L gene; C12L gene; C13L gene; C14L gene; C15L gene; C16L gene; C17L gene; C18L gene; C19L gene; C1L gene; C2L gene; C3L gene; C4L gene; C5R gene; C6L gene; C7R gene; C8L gene; C9L gene; D10L gene; D11L gene; D12L gene; D13L gene; D14L gene; D1L gene; D2L gene; D3L gene; D4L gene; D5L gene; D6L gene; D7L gene; D8L gene; D9L gene; G1L gene; G2L gene; G3L gene; G4L gene; G5L gene; G6L gene; G7L gene; G8L gene; M1L gene; M2L gene; M3L gene; M4L gene; M5L gene; M6R gene; M7L gene; O1L gene; O2L gene; P1L gene; P2L gene.

SOURCE

ORGANISM

Cowpox virus.
Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae; Orthopoxvirus.

REFERENCE

AUTHORS

1 (bases 1 to 52283)
Shchelkunov,S.N., Safronov,P.F., Totmenin,A.V., Petrov,N.A., Ryazankina,O.I., Gutorov,V.V. and Kotwal,G.J.

TITLE

The genomic sequence analysis of the left and right species-specific terminal region of a cowpox virus strain reveals unique sequences and a cluster of intact ORFs for immunomodulatory and host range proteins

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (13-MAR-1997) S.N. Shchelkunov, Institute of Molecular Biology, 633159, Novosibirsk region, Koltsovo, SRCVB 'Vector', RUSSIA
Gene name designation according to Vaccinia virus strain Copenhagen nomenclature, Virology 179:247-266(1990).

FEATURES

source

1..52283
/organism="Cowpox virus"
/strain="GRI-90"
/db_xref="taxon:10243"
complement(1393..2160)
/gene="D1L"
complement(1393..2160)
/gene="D1L"
/codon_start=1
/protein_id="CAA72577.1"
/db_xref="GI:3096963"
/db_xref="SPTREMBL:O73568"

gene

CDS

/translation="MKQYIVLACMLAAAMPASLQOSSSCTEENKHHWGDVILK VTKDQTPNDKICQSVTEITESDPDESSDSTVEDVDPPTTYSIGGGLR MNGFTKCPQIKSISBSADGNTNARLSSVPGQKSDSPAITHEEALAMIKDCEVSD IRCSEKSDIRKTHPVLGSIHSHKVSIEDIIGSTIVDTKCVKNLEFSVRIGDMCKE SSELEVKDGFYVDGASASEGATDQTSIDSTLTKACV"
complement(2286..3341)
/gene="D2L"
complement(2286..3341)
/gene="D2L"
/codon_start=1
/protein_id="CAA72578.1"
/db_xref="GI:3096964"
/db_xref="SPTREMBL:O73559"

gene

CDS

/translation="MKSVLVSYILFSLCIIINGRDIAHPNSGKCKDNRYNRHNLCC LSCPPGTASRLCDSTNTNQTQCGSGFTTFRNNHLPACLSGCRDSCNQVETRSC NTHNRICECAPGYCLLKGSGCKACVSQTGCGIGYGVSGHSTGVDVCSGCLGTY SHVSSADKCEPVSNFTFNYIDVEINLYPVNDTSCRTTITGLSEISTSELTITNNH KCDQPDRCQYFSVLANKVATSGFTGENRYONISKVCTLNFETKCNKNDSSSKQLPKT KNDGGINPHSETVTLVGDCLSSVDIILYSNTNTQDETDTTISYHGVNLVDVSHMPG SCDIHLKITNSKTRFL"
complement(3420..5180)
/gene="D3L"
complement(3420..5180)
/gene="D3L"
/codon_start=1
/protein_id="CAA72579.1"
/db_xref="GI:3096965"
/db_xref="SPTREMBL:O73560"

gene

CDS

/translation="MTAIFYVQGLNHEKEKELCAEHRYDTASNWHDLRKSGKFCDDITF HTKDNKDISAKHVAVAGFSTYFADMLNNDPQKIFNDISGVSLEMAIDFIYTGVMDDLEEDVYEMFDLARELDSEILGATVRSFHYMDERNALMEYLFENRFEKEMEDTYD VMTMEENIANEITELYKDVRFKOLISADSMFDMNCQDLDVKDEADVIDILLDLWTEF PEDKDLIRVIRWNVSDQVREKCKMNPFLFNDVDIDRLSKCYIDKRNARIGFSRY"
complement(10160..12145)
/gene="D8L"
complement(10160..12145)
/gene="D8L"
/codon_start=1

/translation="MDMEDEIVRIVDSMMYIPNFMDDGENEGHISVNNVCHMYLAF /translshfklvikhcdlnkrlkcgnsplhcytmtntfkpSVLKLILHGHSMRNFDSK DEGVHPLHVLHLSIDNKIFDILDPIDDDSKSSDLLLCYLRYKFNGLSNVYVLY KLITKSDPNCVDEDEGLTSLHYCKKHJLSTHESNYKSSHTKRAERFIYTIIDHGA NINAVTIGNPTLHTYLOQYTKHSPRVYALLSRGADTRIRNNJDCPTIMEYINKOCHA TGHILMLNHEQYKGLQKEGOHLLYLFKHNQGYGSHTNILRYLLDREDAQYV EYNTWTFPLHTAFQCNMNVASVLYIGYDINLPTKDDKTQFDFLAFVFNRIIYKADV IINLHHRKLVSLPMKLSFLKMFSEFSPYDDYVYKKIAYCLLRDESEAEHLHSKFLNE DYKSVPMKNISFDKIDSIIIEKCSRDISRLKEIRISDQDLATVLTEDIRVRYLEATH SDRHSFPMYDDLLIEQCHLSMEHKSCLVDKALNKLESTIGQSRLSYLPLPEIMRNITIK LSDIHLNSMLYIGKNHYIYIP
complement(5392..7410)
/gene="D4L"
complement(5392..7410)
/gene="D4L"
/codon_start=1
/protein_id="CAA72580.1"
/db_xref="GI:3096966"
/db_xref="SPTREMBL:O73579"
/translation="MDYCGRECTILRRLLDEEDVYKKIKLEIETCHNLKSHIDRRGN ALHCYVSNKCDTDIKIVRLLSRGVERLRNNEGLTPLGAYSKHRVKYQIVHLLISS YSNSSNELKSNINDFDLYSYMSSDNIDRLLLKYLIVDKRIRPSKNTWAINGLGLVDI IYTPNRPREVLLWLLAKSECTGYVERTCMYOSDMCKNSLHYILSHRESQSLSKDV IKCLIDNNVSIHGRDEGSLPIQYWSCSTIDIEVKLLIKDVTCKYDDISQPIYR GVLAADYLNKRFRVTPYNVDMIEIVNLLIEGHTLIDVMSITSYDSREYNHYIIDNLLK RFOQVDESIVQAMLINLHYGDMVSIPIVRCLMDNGATMDKTDNTYPLIDVFDILNSL YDYNVRFIVENNGHMAVNHVSNRCLMCGYLILSRFNNGYHCYETLIDVFDILSK XMDIDIMDNKTLTYAYDVNNIOFAKRLLEYGANVATSLIINTAOKSSYRE NKKLDVLLLSYHPTLETWIDAFNRDIRYLYPEPLACIRALYLLDDDFSKVAYDIS GRKELKRYRADINRMKNAYISGVSMFDILFKRSKRHLRYAKNPTFINFVSNIKWY KELTSITETVKNGERIDSDVINNTDNLISKLPMEIQREILYYSIK"
complement(7652..8113)
/gene="D5L"
complement(7652..8113)
/gene="D5L"
/codon_start=1
/protein_id="CAA72581.1"
/db_xref="GI:3096967"
/db_xref="SPTREMBL:O73582"
/translation="MASPCAQFRPCHCHATKDSLNTVADVPHCLTEYILNVSHRWTHR ESAGSLYRLILSPRTDATELFGSELKESDSLPWDNIDNCVEIILKCFIRNDSTAYE LRAIIGLQTSIAIVSGRVFNDKYIDILLMLRKILNENDYLTLLDHIHTAKY"
complement(8281..8940)
/gene="D6L"
complement(8281..8940)
/gene="D6L"
/codon_start=1
/protein_id="CAA72582.1"
/db_xref="GI:3096968"
/db_xref="SPTREMBL:O72729"
/translation="MDDKISNAIKAMNNOEVRNIIKNLDQENSTVNDVITNTCNFMR DGVLTKLFFQEFFIKYMPFSVIVAKTVQVGFYNLVKEYLRIVRNLNIDDMGPVKEFTI YEDLAIERYKGPFIIRACVTIVQVEDADFSDSDSDMDDFYNIDLILKLRKSD GTYTVDPVPPTKTAIGTASIIISKTTSINNILDGYPVFGWNIISPSYVNAFLSYMETKV S"
complement(9217..10038)
/gene="D7L"
complement(9217..10038)
/gene="D7L"
/codon_start=1
/protein_id="CAA72583.1"
/db_xref="GI:3096969"
/db_xref="SPTREMBL:O72730"
/translation="MTAIFYVQGLNHEKEKELCAEHRYDTASNWHDLRKSGKFCDDITF HTKDNKDISAKHVAVAGFSTYFADMLNNDPQKIFNDISGVSLEMAIDFIYTGVMDDLEEDVYEMFDLARELDSEILGATVRSFHYMDERNALMEYLFENRFEKEMEDTYD VMTMEENIANEITELYKDVRFKOLISADSMFDMNCQDLDVKDEADVIDILLDLWTEF PEDKDLIRVIRWNVSDQVREKCKMNPFLFNDVDIDRLSKCYIDKRNARIGFSRY"
complement(10160..12145)
/gene="D8L"
complement(10160..12145)
/gene="D8L"
/codon_start=1

/protein_id="CAA72584.1"
/db_xref="GI:3096970"
/db_xref="SPTREMBL:O93130"
/translation="MLYFVEMERYKSTKDASTOTWEETNVVDDRYTHDSITLGYSD
RKNDLWNIYKDLFANGYDLTVKDYNLNTAVSAFYKRDIMNLEMLIDINDLYLWKQRS
SLVHYLDHKKFDYFPFKLVLTNKKCLTYCYCGDFKGFARIFGTPHLVASLKSLS
ITPMYKLLVNGNDINAMNENTPLGTPHLYLHGFTRSNFLDYNETIIDFIELGAD
IIMPANRRIPVITAIHINNIDNIQIKILRKLILSKAKPELFRDFVHMHDYIKKS
YIDFECIDIRSLDGFINDKYDGYTPLHVAIKYNTIRIAKYLDDGADMTLRTYNNK
TIFYLARVYSARKWTSEFLISOLPPELGLPTDHIIDCVLMNKNRVDMWEGSPVVISY
LILWPSFYSKFEAEIEFKIQORDDILVRTEDEDEDEYRYMNNIYYDAEDNPLD
DVNREKINVLHNAIRYISDMVLSKSIYITDGIITLIDLLNNNIKIPVYGNKELIT
ESNYIHIDVXDVTSIGIREKNLMRKVIYYIDHDQOTYLTNPKELVMEIFSNMNLPEI
RDYTKYATNTAIDYIGCCDIDYSLSQMSKSVQOTQTHE"
/gene="D9L"
complement(12159..12386)
/codon_start=1
/protein_id="CAA72585.1"
/db_xref="GI:3096971"
/db_xref="SPTREMBL:O72731"
/translation="MTHIKIRRYKGTFDYVGLYRTNSHPWLNCDKNPYNNIPIRG
VEDYDLNNGISGARIVYADRRCWICEVPMVDY"
/gene="D10L"
complement(12346..12636)
/codon_start=1
/protein_id="CAA72586.1"
/db_xref="GI:3096972"
/db_xref="SPTREMBL:P87596"
/translation="MFTMLRVLVRVIVILSLLSLFLVICSMEHRCFKEETRYKIC
PKNTISINNICYVFSEDTTNWTASFEYCKNMNGRLACIINDYTQDTQIQVI"
/gene="D11L"
complement(12699..14264)
/codon_start=1
/protein_id="CAA72587.1"
/db_xref="GI:3096973"
/db_xref="SPTREMBL:P87597"
/translation="MEERVNIILGQVLTSDKKLLCSYFFKNRFEFDDNTDNLNLSY
DNHDLIKKLINFTELDEDEDQVLMVLADSLNITKAIDLCTEFIERLDAINCL
DIITYAFLNLEDKIAAINTATNETHVVKDDTFLSLPDIISIPENDDLATKSD
EAMDALCINWDDAERINMVKENISKILRVDELLSKYILNELGCENALPPSPVKRSS
GVLVAGISDPDDTYSKLEMYNRVENWKIVSDIPKAINCMAAGIDIVIWCCKQ
TYNFTTNVYSYNTKTNKKPESTVYPRANAGLAVNDYIYILIGTSEPHSDTDS
ARTDPDCGYDNRRYINTPIERRPEDGKWRVKMTMPVAGICSATSIGDNIIIII
NSYKGVYIRCDTKLWRKCIKLSYATNTSAITAYNGIYCIINGFKNIEYRDPTK
SYNIIEKLPRYRHHSLAVLGDELFIYGVGDIINTNATHNVVDVYNHKTNRIRLQKNKL
LSLERPFTVPVRK"
/gene="D12L"
complement(14342..14950)
/codon_start=1
/protein_id="CAA72588.1"
/db_xref="GI:3096974"
/db_xref="SPTREMBL:P87598"
/translation="MINININTILIFASLFAVSFANDYPPPPFFEDKYYITDTPNYISL
DFELYPNVSSCRLTKQSSDVIITSELTITVNSDCDFVTEYYSVKGKTAIAGL
FTDTTKONTSKMCTLNIEVKNAETEPVLIGNFTVRPEKASTHAENFTLIGNCLSDL
HLVYAVNTDEGEEDTATIHGNIIDISYIPNMCATRTIN"
/gene="D13L"
complement(14947..15282)
/codon_start=1
/protein_id="CAA72589.1"
/db_xref="GI:3096975"
/db_xref="SPTREMBL:P87599"

/translation="MKMNTIFLSAIVTCLVYTSFGKTCPNDYILPEPDGLCTACVTCI
SNMVEIQCPGDKPRKCCQCGPLKCTPVAVNSCARCTPDTTIKKIEPTDQCCTTDPNT
KLCYHKYSS"
complement(15374..17668)
/gene="D14L"
complement(15374..17668)
/codon_start=1
/protein_id="CAA72590.1"
/db_xref="GI:3096976"
/db_xref="SPTREMBL:P87600"
/translation="MASLTHEAIVNRKLIYSTYFDFDNKRISTNARINNDMPDGYC
SDGDIYSYNTVRARIKVFVKYICIDNRQEQYTDSTALINIIGSMTLMKADRK
NKSIVDQYKFKVKCIIDKNKSTTNHVFENPNNGMDILTYTFNSPRTGCKLDOIYK
MVDVGVNLNVCKKCTCGILHAYLGNMVDVIDILEMCLNNGVNVNQRLITPLHT
YMTGNVCVHVIKKIIELGGDMCKCYNGTSPIMTYIANTDNIINPELTNLYIESLDGN
KYNIPMMLHSYITLARNIDISYMSFLOPGVKLHKYKDSAGRTCLHQVILRHNIISTSI
IKLHHEYNDINEPDNIGNTVLHYSLSVSVHILDPETDNDIRUQVLCILSLGADI
TAVNCLGYTLPTSICYTAQNYMYDIIDCLISKNVLNKNRILODLLIRVDDTFCII
HHIVAKYNIPTDLYDEPEYDNIIDHVVHCAIIRYNNAVCTSGMTPLHRSYIISH
TNANIVMSFMYLLSIQTNINPTKNGVDPMLTMENMNLGSHOMLVKNILDKRPNV
NMVSFLDKCYTYTCKFPSSLCLCEHDIKPTLRALMALAGLDYCNKCLEYMERDIAID
NSHAMFLAFKLYSIRDNIKLIKLNRSNTNISIYDILVGKCKEKEDLIAIYRHNHLV
ACCHGNDPLDIIINTYITDARSMTYIANDISRYIMDMYIPVMRIPVTLLFSCIIGFRL
TYFKKTIIFNRHDSFINARLTDEE"
/gene="C1L"
complement(18018..19331)
/codon_start=1
/protein_id="CAA72591.1"
/db_xref="GI:3096977"
/db_xref="SPTREMBL:P87601"
/translation="MESADEFMAVDEQFDDLDLWSLSVDDYVKHGLGVDCYVLEPVV
DRKIFRFLLEPCDPVDVLYDFRIARDNIQYIVDRLFAITYKDVISALVSKNTM
EDIFSIIKNCNSVDULLCYLSNAYVIDIVDFMVDHGAIVYIKELCNAYFRGMCKK
ESSVVEFILNCSPDENVDKLDLYKIIQYTRGFLVDEPTVLEICIKPIYEDINQOLD
AGGRTLLYRAIYAGYIDLWSLLENGANVANVMSDGYTCLDVAVDGVSVIARREAHLK
ILEILREPLSIDICKLAILNNTTENHDVTKLCIKYEMVVDYSLCNYVASSLSFDYIID
KCOELEYIROMKIHTMTYELIYNROKNKASHILHRYSKHPVLTCQITCKFGKIYTEV
NEHVTKALNRALIDEIINNVSTDDNLSKLPLEIRLIVSQAIL"
/gene="C2L"
complement(19440..19976)
/codon_start=1
/protein_id="CAA72552.1"
/db_xref="GI:3096978"
/db_xref="SPTREMBL:P87602"
/translation="MDSRIAYVLVSASLLSVNCHKLVHYFNLIKINGSDIITNADIL
LDNYPMTFGDKDIYPSAFMGVGNKLFLDLYKNIFEFFRLFRVSVSSQVEELEYYYS
CDYTNNRPTIKOHFYNGEYVEIDRSKKAANKNSWLITSGFRLQKQWFSDEDCIILR
SLVRNEDSNKSKSLST"
complement(20088..22589)
/gene="C3L"
complement(20088..22589)
/codon_start=1
/protein_id="CAA72553.1"
/db_xref="GI:3096979"
/db_xref="SPTREMBL:P87603"
/translation="MNSLIENSVLHVRKLLYTHFNIDIDHAPTATSRRDCENQYDSYD
IYVRSVFNKLNLVLEYRDIYVYKLINSTISHTINYDQTYDNDIKLSLEATTFRRHRY
ISSSRKKSIDSIEOMYSOI IKDITTELINDENVINVDNCLCIDYDDNDKDDKDD
KDNKKDDKNKDDKDDDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD
NSDKYSSNFSYVYFKAHTIGIDIKLVKWMITGVDINYKNYSGSLIHTYILGNMYVD
TDVLELLCRSGKNINEONNHLITPLHSLRREDSISASVLKKVIEIGADNRCCMKL
TPIMTYIINYAMNIDBEITIKYMDNDKNITNIPELIYIQMAQYVDLDVIAKHLNNSV
ILDYKDSVGRCTLHNTLMRNYINIDVYVKDLLEHNGVLTNETLNGTNYHTYLSRLCIV
NKNLTVDSSINNNTIEVVKILRSGDLSLVNNGYTPLTYSYCTAONTMYHDIIDC
LMSDKVLNMYKHRIIODLITREDISLPIIHHVTIKYINPVDSYTDVEYCNVETNDY
HRQIMYKNNKIYVFPFGMTAFHVAVCLRKDAMKYLISIGSIDKPTIKGTALMLFPH
RDRISNCHIVKMLDNRPSIDNIIIFLDKCYKNESLUSLSCSTKNTSILEMWSAFMLI

gene

complement(33086. .33703)

CDS

/gene="C15L"
complement(33086. .33703)
/gene="C15L"
/codon_start=1
/protein_id="CAA72565.1"
/db_xref="GI:3096991"
/db_xref="SPTREMBL:P87614"

/translation="MDTYMRLDLDKLUKHENIFSDNIIEDEAKETFGSRKIYDTSVDD
LIELYSAKYLNQNLKDVVIERMDYCKYIGKDNWSTIYSFYKENGIRLNSFLRQYIN
NNIEICNTDQFLKSDVCDILDNDIEIVVTRYETILNMLVLRLENKRNVNDOTFKV
MFVIRFKFITVSELTNAIEKIAPEYRQCLQDLXHKITRPHEDN"
complement(33774. .34721)
/gene="C16L"
complement(33774. .34721)
/gene="C16L"
/codon_start=1
/protein_id="CAA72566.1"
/db_xref="GI:3096992"
/db_xref="SPTREMBL:P87615"

gene

complement(33774. .34721)

CDS

/translation="MDTIKIFNHGEFDTIRNELVNLKVVKNWNTINSVTSDDTID
ISDCIRILYKQFNVRNIEVSSDISFIKYRFNDTTLTDNVGYLVIYLNRTKSVKT
LIYPTSPVITSSIEDIMFSLNFRFENVRKDYKLVKMSISLTYKPSICRIQYDNKY
IDISDSQECNNICYVITMDPHLIDLETICVLVDKSGKCLLVNEFYIRFRKNHIYNS
FADLCMDHIFELPNTKELFTLRDNDGRIANDNDKLESGNNTWIPKTDDEYKFLSKLM
NIAKFNTKFDYVVLVGDTDECTVFTFKVKYINLNYE"
complement(34788. .35567)
/gene="C17L"
complement(34788. .35567)
/gene="C17L"
/codon_start=1
/protein_id="CAA72567.1"
/db_xref="GI:3096993"
/db_xref="SPTREMBL:P87616"

gene

complement(34788. .35567)

CDS

/translation="MKVESYFTLTLTGIGCVLSCCPIPSRPTMKFKGTVDSHVNI
TIYCLPGYRQKQMPIYAKCTGTWTLNQCIRKRCPSRPTDINGQLDGGVDFGS
SITYSCNSGYHLIGESKYSCELGSTGSMVWNPAPICESVKQSPSISNGRINGYED
FYTDGVSVYTCNSGYSLIGNSGVLCSGGSEWSPPTCQIVKCPHTISNGYLSGFKR
SYSYNDNVDFCKYGYKLSGSSSTSCSPGNTWQFPLKQVR"
complement(35634. .37172)
/gene="C18L"
complement(35634. .37172)
/gene="C18L"
/codon_start=1
/protein_id="CAA72568.1"
/db_xref="GI:3096994"
/db_xref="SPTREMBL:P87617"

gene

complement(35634. .37172)

CDS

/translation="MESVIFSGEIIQVKNKEIITASPYNFFKRIODHHLKDEAILIN
GINVHAFESLDYMRWKINITINNVMILVAAI IIDVPVVDLCVKMTIHNINSTNC
IRMFNFSKRYGIKKLYNASSEIINNITAVTSDPEFKLSKDELTI LSHEDYVNRHE
DVTAMILLKWIHKNDVDIILNHPKFMWNRNALSILGLTISKSTKPTVRNGIH
NIVYKNSDIYSTITHYSRPTXYTWIGNTDROFYANVLHNCLYITIGGINNRHYYS
VSRVDLTKRKWKTVMNSLSKSEVTCVNDGKLYVIGGLEFSISTGVAEYLKHGTSKW
IRLPLNITPRYSGASVFNDDIYVMGVYTYTEKYVVLNDVECFTKNRWIKKSPMPRH
HSIVYAVEYDGIYAITGITHTERNLYKYIVKEDKWIELYMYFNHVGKWFVCSGCDY
ILLIADAKYEYEPKSNWNLFDNSTRIEYDMFTKDEFPKCNVTHKSLPSFLSNCEK
QFLQ"

gene

complement(37242. .37937)

CDS

/gene="C19L"
complement(37242. .37937)
/gene="C19L"
/codon_start=1
/protein_id="CAA72569.1"
/db_xref="GI:3096995"
/db_xref="SPTREMBL:P87618"

gene

complement(37924. .38277)

CDS

/translation="MVKSNKIQKNISNSCCOYIMWNTNNILMRHLKNLTDDEPKC
IITHRSSDLYLSDRDVTSITKETLVSEIVEEYPPDCNKILAIFVLVDKXIDVDIETK
LKPKPRAVAILDKKTADIKLIDLVRHYERYIEODIPLGLFKAKIDSYRTRINKYSK
EGLGATEYFNKXHLMFYTLPIPNYRFFCRNSIGFLAVLSPTIGHVKAFYKFEIYVSI
DORRFPKKEMLSK"
complement(37924. .38277)
/gene="P1L"
complement(37924. .38277)

/gene="P1L"
/codon_start=1
/protein_id="CAA72570.1"
/db_xref="GI:3096996"
/db_xref="SPTREMBL:P87619"

/translation="MRTILIRIILWRNDGDFSYNDCKFKMLMLDELVDGDDVCTFLIK
NMRMTLSDGPLDLRLNQPNVNEVDAKRMIAISAKVARDIGERSEIRWEESETILFRMI
EKYFDLMTDLYGEK"
complement(38413. .38940)
/gene="P2L"
complement(38413. .38940)
/gene="P2L"
/codon_start=1
/protein_id="CAA72571.1"
/db_xref="GI:3096997"
/db_xref="SPTREMBL:P87620"

/translation="MTSAMDNNEPKVLEWYDAPILPEGSSMDPNIIDCINRHNMC
IQRTYSSSIALLDRFLMKNDELNTQCHILKEFTYEQMAIDHYGVVAILYQVIR
KRPQNHHTIDLFKRIKRTYDFTKVDPVEFVKVIGFVSIILNKYPVSYLVYENLVY
DEFKCFDIYVETKYF"
complement(38982. .40406)
/gene="O1L"
complement(38982. .40406)
/gene="O1L"
/codon_start=1
/protein_id="CAA72572.1"
/db_xref="GI:3096998"
/db_xref="SPTREMBL:P87621"

/translation="MIEVIESKLLQIYRNRRNRNINFYTTIDNIMSAEYVLSLYAKY
NSKNLDVFRNMQAIEFSGNNYHILHAYCGIKGLDERFVEELHIGYSPNETDDGNY
PLHIASKINNRIIVATLLAHGADNACDKQHTPLYLSGTFDDEVIERINLLRYGAK
INNSVDEEGCPLLACTDPSERVEFKIMSGFEARIVDFKGNHHRHLMNDNPKAST
ISWMLKGISPSKPDHGNTPHIVCSKTVKNVDIIDLPLSTDVNKONKFGDSDPLTL
LIKTLPAHLINKLLSTNVITDQTVNICIFYDRDDVLEIILNDKGYQVDSDFKMAVE
VGSTRCVKYLNDNDIICEDAMWYAVLSEYETWVDYLLFNHESVDSVYNGHFCMSECV
LNPNVILSKLMLNDPSETMYLNMKAIEKKIDKSIILPIFYFLVLMHPDFCKNRRYF
TSKRFYDIDVHGEVSYEVDDDYF"
complement(40384. .40875)
/gene="O2L"
complement(40384. .40875)
/gene="O2L"
/codon_start=1
/protein_id="CAA72573.1"
/db_xref="GI:3096999"
/db_xref="SPTREMBL:P87622"

/translation="MSBSYIDRNANIVLTGVLGINMTIMDORFVAAAGVGVKDNK
LSVLLFTTQRLDKVHHNIVSITTCMMNCGTKYDSDLPESIHKSSCDITINGSCVT
CVNLDTPTKINPHYLHPKDKLYHNSEYGMRSYGVTTFIDELNQCLLDIKELSYDIC
YRE"
complement(41155. .42009)
/gene="M1L"
complement(41155. .42009)
/gene="M1L"
/codon_start=1
/protein_id="CAA72574.1"
/db_xref="GI:3097000"
/db_xref="SPTREMBL:P87623"

/translation="MDLSRINTWKSQOLKSLSSKDAFKADVHGHSAIYVAIADNNVR
LVCTLLNAGALKNLLENFPLHOAATLEDTPKIAKILLFSGMDDSOQFDKDKNTALIYAV
DSGNMQTVKLFVYKKNRMLFYGTGKTSTFYHVMNLNDVSIYSTFLSEIPTFDLAIL
YSCIHDSIRKNGVNDMMILLDDIWMESTNNLSLFPIDIKLAIDNKDLEMLQALFKYDI
NIYSANLENVLLDDAEITAKMIIIEKHVEYKSDSYTKDLDIVKNKLDLEIISKNKELRLM
YVNCVKKN"
complement(42228. .43349)
/gene="M2L"
complement(42228. .43349)
/gene="M2L"
/codon_start=1
/protein_id="CAA72575.1"
/db_xref="GI:3097001"
/db_xref="SPTREMBL:P87624"

/translation="MIALLILSACSASAYRLOGFTNAGIVAYKNIQDGNEDDNIVFS
PFGYSFMPMSLLPASGNTRIIELLKTMKRDGLGPAFTELLISGLAKLRTSKYTYDL

gene

complement(42228. .43349)

CDS

/gene="M2L"
complement(42228. .43349)
/gene="M2L"
/codon_start=1
/protein_id="CAA72575.1"
/db_xref="GI:3097001"
/db_xref="SPTREMBL:P87624"

gene

complement(37924. .38277)

CDS

complement(37924. .38277)

TYQSFVDNVCIKPSPYQOYHRRGLYRLNRRDAVNKINSIVERSGMSNVVDSNMLD
NNTLWAIINTIYFKGTYOYFDDITKTRNASFTNKYGTKTVPMMNVVTKLQGNITIDD
EYDWMYRLPYKOANISMYLAIGDNMTHTFDSITAAKLDYWSSQLGNKYVNLKLPRESI
ENKROTKSIAEMMAPSMFENPDNASPKHMTROPPLYIKMFONAKIDVDEQGTVAEASTI
MVARSSPEKELEFNTPFVFIIRHIDITGITLPMGKVESP"
/gene="M3L"
complement(43400. .43666)
/gene="M3L"
complement(43400. .43666)
/gene="M3L"
/codon_start=1
/protein_id="CAA72576.1"
/db_xref="GI:3097002"
/db_xref="SPTREMBL:P87625"
/translation="MLAFCYSLPNAGDVIKGRVYENDYALYIYLFDPHSEAILAESV
KMHMDRYEYRDKLVGKTVKVRVYRDYTKGYIDVNYKRCRHQ"
/gene="M4L"
complement(43726. .45000)
/gene="M4L"
complement(43726. .45000)
/codon_start=1
/protein_id="CAA72540.1"
/db_xref="GI:3097003"
/db_xref="SPTREMBL:P87626"
/translation="MNPDNNTIAVITETIPIGMQFDKVLSTFNMWREILSNTTKTLDI
SBFYSLSDVETNGFTIILNELIQLPKRGVRVRVAVNKSNKPLKDVRLQWAGVEVR
YIDTINILGVLHTRFWISDNTHIYLGSAWMDWRSLTQVKELGIAIFNNRNLAADLTQ
IEFVYWLGVNLPYWNKFPAYINTDHPLSMNVSGVPHSVFIASAPQOOLCTMERTN
DITALLSCIRNAKSFVYVSMNFIPIIYKAGNILFPVYIEDELRSADROVSVKLL
ISCWORSFIMBNFLRSIAMLSKNIDIEVKLFIVPDADPPPIPYSRVNAHKYAVTDKT
AIGTSGNWTGNFTDFCGASINITPDGUGLQKQOLEDIFIRDWSNXYELIYDTSPTK
RCRLKMLNKQCTNDIYCDIEQKEIPEYSLE"
/gene="M5L"
complement(45027. .45857)
/gene="M5L"
complement(45027. .45857)
/codon_start=1
/protein_id="CAA72541.1"
/db_xref="GI:3097004"
/db_xref="SPTREMBL:P87627"
/translation="MSANCMENLDNDYIYCKYWKPIYPKALVFIISHGAGEHSGRYDE
LAENISSLGILVFSHDHIGHSNGCEKMMIDDEGYVVRDVQHVVTIKSTYFGVPVFL
LGHSMGATISILAAENPDLFRAMILMSPLVNADAVPRNLNLLAAKUMGTITPNAPVGK
LCPESVSDMDEVYKYQDPLYNHEKIKAGFASQVLKATNKVKRIIPKINTPTLILOQ
TNEISDSVAGYEFYMOHANCNREIKIYEGAKHLHKETDEVKSKVMKEITWIFNRVK
"
/gene="M6R"
45960. .46445
/gene="M6R"
/gene="M6R"
/codon_start=1
/protein_id="CAA72543.1"
/db_xref="GI:3097006"
/db_xref="SPTREMBL:P87628"
/translation="MKETDTVFCEIVMATKSDYEDAVFVVDDETCSDRSDSIDLIDE
YITWRNHVIFNRKDTISGRGLYKELMKFDAAARIYIGDKINEIVPAMSEGDIYNLT
EYHDQESLFATIGCAIKETHWGYKKISESRFQSLGNTIDLMTDDNINILILFLEKKL
N"
complement(46277. .46471)
/gene="M7L"
complement(46277. .46471)
/gene="M7L"
/codon_start=1
/protein_id="CAA72542.1"
/db_xref="GI:3097005"
/db_xref="SPTREMBL:O72732"
/translation="MRYEDPLYHQFNFFSRKSKTKMFILLSVIRSVNMFPPNDWNLDSEIF
LYPQCSVILAHIPMVANRLS"
complement(46512. .47228)
/gene="G1L"
complement(46512. .47228)
/gene="G1L"
complement(46512. .47228)
/codon_start=1
/protein_id="CAA72544.1"

/db_xref="GI:3097007"
/db_xref="SPTREMBL:P87629"
/translation="MDNSMLSMEMCNINVDYVDEDETHVDIDNGIVHHIEDEASDND
DHQYVPLPENMVYRKSTNILDYSTERDHYVMAVOYVMSKORLDDLYROLPTKTR
SYDIINIYCDKVSNDYNRDMNTIMYDMASTEFTYVDINNEVNTILMDNKGGLVRLAT
ISFITELGRRCMNPVETIKMFTLLSHTICDDYFVDYITDISTPRDNATNTSSSTREYLK
LMGIAMIFATYKTLKYMIG"
complement(47228. .47671)
/gene="G2L"
complement(47228. .47671)
/gene="G2L"
/codon_start=1
/protein_id="CAA72545.1"
/db_xref="GI:3097008"
/db_xref="SPTREMBL:P87630"
/translation="MFNNMINSVPFVKETNRKASPTPROSPAAGYDLYXSAYDYTIPTP
GERQLIKTDISMPKFCYGRAPRSGLSLKIGDIGGVDEYDRGNIGVILLNNGKC
TFNQITGDRIAQLIYQRIIYPELEEVQSLDSTDRGAQGFGSTGLR"
/gene="G3L"
complement(47695. .49152)
/gene="G3L"
complement(47695. .49152)
/codon_start=1
/protein_id="CAA72546.1"
/db_xref="GI:3097009"
/db_xref="SPTREMBL:P87631"
/translation="MPFIVNTVYCKNILALSMTKKFKTIIEAIGGNILVNSTILKKLS
PYFTRHLROKYTKNDPQTRVCLDDIHSITSIVIYSTGKVYIDRHNVINLLRASIL
TSVEFIYIYICNIFILDRFRKEYCECYMMGIEGLSNLLCHTKDFTAKHFLLEDII
DNFDYLLMLLIESLDELPDEYVYDFVIKWTMRNRNLGNLLILKIKVIRSNYLSIP
RGINNYKWLIDCTKIFHCDQPKRSYKPIFYEYPMNNDQIIDIFHMCSTHGVSVTVE
IGWMNNEIHNAIAFVNIYSNNWPIPPMNSPRLYASGPANNNKLVXVGGGLPNPTVSE
RWFHGDAAWNMPSLKPRCPAVASINNVYVMGGHSETDTTTEYLLPNHMQWQFGP
STYYPHYKSCALVFGHRLFLVGNAEFYCESSNTWTLLIDDPYIPRNPPELLIIVDKLL
LIGGFYRGSYIDIEYNNRTYSWNITWDGNEWGCK"
complement(49163. .50122)
/gene="G4L"
complement(49163. .50122)
/gene="G4L"
/codon_start=1
/protein_id="CAA72547.1"
/db_xref="GI:3097010"
/db_xref="SPTREMBL:P87632"
/translation="MEPI LAPNENRFEVFIPIQYHDIWNMYKKAESFWTVEEVDISKD
INDWNKLTDPDEKYFIKHVLAFFAASGIYENLAERFCTEVOITEARCFYGFOMAJEN
IHSEWYSLIIDTVYKDSNEKNYLFNAETMPCVKKKADMAQKWIHDSAGYGERLIJAF
AVEGIFSGSFASIPFWLKKRDLMPGLTFNELLISRDGLHCDPACLMFKHLLHPSEE
TVKSIITDAVSIHQEFLTAALPVKLLGMNCEMMKTYIEFVADRLLISELGFKKIYNVTN
PFDPMENISLEGKTNPFERKRVGEYQKMGVMSQEDNHFSLDVPF"
complement(50154. .51125)
/gene="G5L"
complement(50154. .51125)
/gene="G5L"
/codon_start=1
/protein_id="CAA72548.1"
/db_xref="GI:3097011"
/db_xref="SPTREMBL:P87633"
/translation="MCTNVVRVFVILYLLAVCGCIEYVDVDDNVHICTHTDVSINHITS
WYNDRVIALATEDKTSGISFIFKRVKNISLTCNLNLISSLRYESGTYKGVSHLKDGYI
YITPMNISVKANIIDTGRVRYLTRNYCEVKIKCEITSFALNGSTTTPPHMILGTVDKWK
KYLPPPTDDYRYVYELKRYISGNPYPTESLALIEISSTFNRFITIVKLNDDFCYCLFS
QNYSPHKMLNARHICESNEALNNDNASSMPSSHNNLANDLSSMMSQLQNDNDNDN
NDYSAPMNVNLMIVLITMLSTILYIVVIAAISWYKRSKYRHIDN"
complement(51155. .51379)
/gene="G6L"
complement(51155. .51379)
/gene="G6L"
/codon_start=1
/protein_id="CAA72549.1"
/db_xref="GI:3097012"
/translation="MSKILITFVKNKIIDLINNDQIKYSRVIMIEESDSLLPVDEVHAN
HGFDVCVEMIDENISNEIQKTESFTTN"
complement(51395. .51637)
/gene="G6L"

gene 16828. .17679
/gene="B11R"
CDS 16828. .17679
/gene="B11R"
/codon_start=1
/protein_id="CAA75287.1"
/db_xref="GI:3097032"
/db_xref="SPTREMBL:O72748"
/translation="MESFKYCFONDGKKWIIIGNTLXSGNSLLYKVRKNFTSSFYNYVM
KIDKSHKPLLSSEIRFYISVLDPITDNWTERIGIKYLAIPDLXGIGETDDYMEFVIK
NSGRVAPKDTESVFEACVTMINTLEFIHSGFETHGRIEPRNIIIRKRLIDDIYFVVK
NKLYSGNSHDYNEEDMITSGNIDYMCIDNHLGATVSRGDLMLGYCMIEWFGKRL
WNKSSIKVYKQKEYKFIATFEDCFPEGNEPLDVRVYIELVYTLIDYSQTPNYDRL
RRLFIQD"
gene 17776. .18813
/gene="B12R"
CDS 17776. .18813
/gene="B12R"
/codon_start=1
/protein_id="CAA75288.1"
/db_xref="GI:3097033"
/db_xref="SPTREMBL:O72749"
/translation="MDIFREIASMKGENVFIISPASISSVLTILYYGASTAEQLSK
YVEKEENTDVSQAQNIISFKSMNKVYGRYSAVFKDSFLGKIGDKFQTVDFDCTIDAI
NKVDYITECKINPLLTLEOLSPDTCILAI SA VYFKAKWLPPEKEFTSDYDFYVSPT
MVDVSMMSYMGELFNHASVESFGNESI IELPYGDTSMVMVILPDKIDGLGIESIQNL
DTNFKWCNSLEATFIDVHIPKFKYGSYNLVDTLVKGLGTDVYFSGDYSNNCNSDV
SVAMIKHYIDVNEEYETAAATSVLADCASTVTNEFCADHPFIYVIRHVDGKILF
VGRYCSPTTNC"
gene 18889. .19338
/gene="B13R"
CDS 18889. .19338
/gene="B13R"
/codon_start=1
/protein_id="CAA75289.1"
/db_xref="GI:3097034"
/db_xref="SPTREMBL:O72750"
/translation="MTANFSTWFSQHCQCDRLTSIDDVROCLTBEYIYWSVAYNRN
OCAGQLYDPLLSRDOAESVFDIRELVKNMPDWDKDCVEIIRICYPDEQKTIRES
AIIIGLCAYATYGGEDHDTSTNSLNFVWLEMLNVYDNIIFRRMN"
gene 19430. .20410
/gene="B14R"
CDS 19430. .20410
/gene="B14R"
/codon_start=1
/protein_id="CAA75290.1"
/db_xref="GI:3097035"
/db_xref="SPTREMBL:O72751"
/translation="MSILFVIFLPIFFYSSVQTFNAPECIDKGOVFAFSEMELENEPV
ILPCQINTPLSSGYNILDILWCKRGADNDRIIPIDNGSNMLILNPQSDSGIYICIT
NETYCDMSNLNITIVSVESNIDLISYQIVNERSTGENWVCNINAFIASVYNADIIW
SGHRRLRNKRLKQRTPGIITIEDVRKNADGYTCVLEYIIRGKTINVTIRIVKLEVRD
IITHPTMQLEGVVTSIGSLNTIACRVSRPPTDADVFWISNGMYEEDDGDGGRIS
VANKIYTDKRRVITSRILKINPKVEEDATTTFTCAFTIPSISKTVTVSIT"
gene complement(20456. .21478)
/gene="B15L"
CDS complement(20456. .21478)
/gene="B15L"
/codon_start=1
/protein_id="CAA75291.1"
/db_xref="GI:3097036"
/db_xref="SPTREMBL:O72752"
/translation="MSRKFMOVIEYDREQYLDDEFTEDRYNDSFTISPEYISAEKYMCR
YTLTHNHCINVRCAOLSKLLDIITNCKIYNKIELVRAFRVYVILDKIKNWSKVG
SHAKVLPVIFTHSTRNDKSVKTYGKIYNNIELKRCADHAIVYDFPKFLTLPNKT
SHAKVLTFCKLKVTITPEAPLPGNVLVYTFPDINKRIPGYIHVNIIEGCDGMIYIN
SKKACVLCMLRSMYRIPFPFIDICSCGQYTNDDIEIPIHDLIKDAIFKNKETVY
LKLNNKTIARETYENNIDTAITQEBHEYVKALGIVCKLMINMHSHIVGVNHSNFTVNC
LLEDNV"
gene 21618. .23342
/gene="B16R"
CDS 21618. .23342
/gene="B16R"

gene 23409. .24464
/gene="B17R"
CDS 23409. .24464
/gene="B17R"
/codon_start=1
/protein_id="CAA75293.1"
/db_xref="GI:3097038"
/db_xref="SPTREMBL:O72754"
/translation="MTMTMMVHIYFVSLSLFLHSAIDAIDIEIETEFTNKMRTDLP
AKDSKLNIPACMFGGTINDMAALGEPESAKPPIEDSLSHRYKDYVVMNEKIGKTRR
PLNRVYKNGDLWIANVTSDNSRRRYLCTVTTKNGDCVQGIVRSHIRKPPSCPITYEL
GTHDKYIDILYCGILYAKHYNNITWYKDNKEINDDIKYSQTKELIHNPELEDSOR
YDCVYHVDVRIKNDIVSRCKILVTIPSODHREKLLIDPKINVTIGEPANITCTVAS
TSLLIDDLLEWENPSPGWIIGLDFGVYSILTSRGITEATLIFENVTVEEYIGNTYTCR
GHNYFDKLLITTTVVLE"
gene 24532. .26919
/gene="B18R"
CDS 24532. .26919
/gene="B18R"
/codon_start=1
/protein_id="CAA75294.1"
/db_xref="GI:3097039"
/db_xref="SPTREMBL:O72755"
/translation="MDEDTRLSLYLTDREHINVDISI KOLCKISDPNACYRCGCTAL
HEYFYNVRVNGKYKYNGYQYVSSDYENYVYDDYDKTGNSSEDSSESIN
ISIKTEYENEYEFYDETQOQTOGVLDYDKLTNEDDFDEFGYDRSVGVHDIIDES
INKVYGRSHVWRDIMEHNDGVYSIGKEYIDNIYEDRHTIDEFKIDSDVDVDA
EHSQIINDVSTQTEWKSSELDYRMEYSIPHRYSKHSVFKSKKNDLNMNVYKE
LISNGASLTIKDSNNKDPJAVFRRTIMNLEDIINKHTTIYERYIYVHTYLNKRN
FDYPPFKVLTKNHLNYYNISDSKYGTPLHLILASNNKLITPNYMKLLVYNGNDIN
ARGEDTQMRTPHLKLYCKFYHNIIEYGRYNEKIIDAFELGADLTIPNDGCMIPVY
YCHSNAEYGYNNITNIIKILNLSRHAPNLFRDVRVMDYISNTYIDLECLDIIR
SLDGEDINGYFEGRTPLHCAIQHNFTQIAEYLLDRGADISLTKDDGKTVPDLSLSYI
PLKWSFLSRLPPKSVICSLTNHIIIDYVLTNNRRIIWSOMINKYVLLDPSFSYRF
RNAISELDQYNNRYNRFEDRDVNEKYGVKVLHDLDTYIKDQVVKLSTSTNNITLY
DPIINNKSEFPERRANDKOLIINLIKSTYHNLIKIEKVTLEKYTLTNIVLEYMISSR
SQSSYLSRIPNEILLEILYKLDWYDLRNLTRYTRENDIAEYHIENTRSTSTQTE"
gene 27021. .28694
/gene="B19R"
CDS 27021. .28694
/gene="B19R"
/codon_start=1
/protein_id="CAA75295.1"
/db_xref="GI:3097040"
/db_xref="SPTREMBL:O72756"
/translation="MDIENDIRNRRIIRNISLDDDDILCDVVIITIGDGEIEKAHKTII
LAAGSTYFKMETTPMIAROLVTRVNLQMFQDAVKNIYQVLYNRHISMMNVDVILVK
ADYLLIDDLVADCESYIKDYINHDTCIYMHKLYEMVHIPYIKRMLSMNIPITLIT
TDAFKKTFEILFDIISTNDNVLYREGYKVTILLKWEYNIHTTEOLLCLISCDIDQ
NLDKRSLLLSNKTINMYPSCIOFLDNKONRNIIPROLCLACHDQVNVYCNPCILY
YNIIMYEVSLTIPNHLINAYASALVDNEIILAGYNNPNSLNKYKINIKENKIHVE
LPPMLKNCRSLAVDITDIYAIGQNGTNSIDMETNISKRYTNTVNNIWTLPNF
GMCVLDQYIYIIGGRTQHDITVSHVTSNIDMETNISKRYTNTVNNIWTLPNF
WTGTLNPGVSHKDIYVVCIDKDEKNYKTCIFRYTNTNTYNGWELVWTTSRSLALHT
ILHDNTIMLHCHYESYEMQLQDTFNVTREWNHMHCHQSNSYIMHNILPIY"
gene 28925. .30052
/gene="B20R"
CDS 28925. .30052

/gene="B20R"
/codon_start=1
/protein_id="CAA75296.1"
/db_xref="GI:3097041"
/db_xref="SPTREMBL:072757"
/translation="MTIIYIRYRYLYVYAMDIFKELILKHPDENVLISPVLSILSL
SILHGAAGSTAQSLSKYIENVAENASEDKDNNDMDYDIPYCATLAPLKIYGS
SIEFHAFLOKIKDDQFQTVAFNANQTKELINEMWTMTNGKINSLSLTSLNTRMT
VSAVHEKAMWKYPFSKHLTYTKFYISKNTVTSVMMVCAENDLOVYHINELFGGFS
IIDIPYEGNASMTIILPDDIEGYNIEKNTDEKIKWRGMLSPKSIDLMPKPKVEM
TEYNLPILENLGLNIFGYADAFKMCNETHIVEKFLHKTIFDIDVNEETEASAVTG
VEMTFNSVTRTKVIINHPPYMIKONTGRILFLGKYCYPO"
30212. .30784
/gene="B21R"
30212. .30784
/gene="B21R"
/codon_start=1
/protein_id="CAA75297.1"
/db_xref="GI:3097042"
/db_xref="SPTREMBL:072758"
/translation="MTIYGLIACILFVTSSTIASPLYIPVIPPISEDKSFNSVEVLVL
FPDDKDYVTSQFNNTIDTKDWIINVLSIPDGLDIPLNITYWSRFTIGRALFKSE
SEDIQKNMSILGVSIECKKPSLLTFLTYRKMTRVFNRPDMAYIRGDCLEALIVYM
TYANKTGETDYIYLSNGGLPAYIRNGVDG"
31045. .36846
/gene="B22R"
31045. .36846
/gene="B22R"
/codon_start=1
/protein_id="CAA75298.1"
/db_xref="GI:3097043"
/db_xref="SPTREMBL:072759"
/translation="MNLQKLSLAIFYLATCSWCYETCMRKALYVDIQLHVEDNKDS
VASLPYKLOVNORESRLLATFNWTGIAESVRNEFIKIDVNGTYLYNKTAVSIL
IDSTEELPTVPTTTEPSTYNTIDISTVITTEEOVTSPPTPYAPVTPILPISAV
PYDORNNVSTISIQVLSKILGVNETELNYLITHKNATVDNNTINNNTSSNTKYD
DEYSDNNLTGNIQFLEINCCYNVSVSDASFRIILVNDTSEEILLMTITSSDFTFS
STNTECKLTINNITSNITVRITQNMNVTSCDKCSMLNLTVIPAVNPFNFTLMKI
GVKDEENVTYNYNCKLTNTSCDELINLDEVINNTLTNIIRNSVSTNSKRRLD
NGEPEFSTKELDCLYESYGVNEDVSHCFASPRRSDDKKEYMDMKLFDAHAKDLRI
DSVIPRGTTHFQVAGSAGGVGDRSPFQNVKSRASLSIAEKIMPRVTTATEQLYA
TVNQTQIPAGVASTPTEALVSTINQKLSAKEVTYASLNLPGSSGYVHRPSDVIY
SSIRLSLPDSDSDEYDIQTVKEYNERYGRPVSRTOSSSESDFEDIDTVREIKQ
QKYGASRVRTSSSSDFEDIDTVREYRQKYGNAMAKGRSSSPKSDPLYSTVKKTQ
SLSGVDIVTKQSDYSLLPDVNTGSSIVSLPTRKGATRRPRPTNDGLOSPNDPPNP
RRPLPDHDDYSPQVPOQDYSPRRPPLLPKQVONPRPLPPRAGOLPIDQPDGCF
SKVSPRRCRASGVICGMIOQKPNDDYSLQRPKLEFEIYAEVNGIPIKNNVPFVIG
NKHSKITYSAMSKLSTKFDKSTAFGAAMLLTGOQAISSQTRSTALSCKDQMSQEEKIF
EAVTMSLTIGSTLTSAGMAGPKLMTAGMAISAITGIIDTKIDYILFSGQRPADP
VIKLFNTYAGILLSNKMVGKCLTPGDDTLIYIAYRNDTSFKHNTDAMSLYFLDVID
SEILYLNLSNLYQLKVACPIGTLRSVDVDITAYTILYDADNIKKYKIFRMATLL
SKHPVIRLTCGLAATLVIKPEYVPSIDMQLLKMATPGEPSTRSIPSDVCDRPLKFF
YLLAGGCPDTSQTFIVHTTCSILLRTATRDQFNRWVLQNPFRQEGEYIKLQTFISKY
DFNDTIIIDPQVAGHASFCTNRSNOCFWSEPMILEDVSCSSRTRKIYVKLGFNAE
YDRFKSDIITVFNENKISGVFWFSMTGSVTSRNMFTLSRGMPPFRSTYCDNRRSG
CYVAGIPFHEDSVPADLHYGPEIMLKETYDNTSIDPRVITKSKTHFPSPISVKFMA
LGNGYDNPSEWEDAKTKKITYSAMTIKVLPCVTVRNNIDFGYNIGDIIISNMYLQST
SQDYGDTGTFKFSRSDHCESSDLTSKCVTVTPAFSPINRSTIEGLCFSVTT
SKCHATSNNGLKSSGKGDVAKVPRACFHHWNYTSLDYSYEDIMWSNMPDYD
KSDYIHEYDRDTWIESNLQOPQPYFEFTHGNSNEYVDKEISKNLDINAYKIMEY
SDGLSPASINRLAKATSEGREIASVINDGNLLDIAVOADKEKMADIOTRINDITROL
FMHTLSDPKIDKDIIESESGKCCIIDVKNRVEKYPIDNYLCGTLDDYIYTSVEYNK
SYVLVNDYMSYDYLESSGIVVLSYCEWNTIISLDTKAKDAIEDIVIVASVAERALNDM
FKFEKNNVSAIILEEDNYLNSSPDIYHIIYIIGGTILLLLLIIILAIYIACNKYET
FKYKIMYDNRISKSEHSDSLSTVSMELIDNRY"
37121. .38866
/gene="K1R"
37121. .38866
/gene="K1R"
/codon_start=1
/protein_id="CAA75299.1"
/db_xref="GI:3097044"

/db_xref="SPTREMBL:072760"
/translation="MLRDYDAKMFNDTVLDFISSEMLESLESIITFTNNNSSTNIDNMC
HLXKYVEVCPSSILPFLFVECCDINKLVEGCTPLHCYLMANEGPESSEYKNLLIHYIYSRDLKPI
NTENVRDSGHIPLHKYLTHTDKVENDLFDMLSDSIDSEFSEYKNLLIHYIYSRDLKPI
NYIYKLLRKAGADPNADDGNTFLHFYFCIYMSAYEKMSFNRKREKKFIKELYKVG
ADINKVNIINTPLUHYVSQDIPRIFTLLSGLADLTQNNDRLTPIMEYIKCEYI
EYNTLPLINWYSEYKRLKEEQHLLHLFIHRNDSNGLNIIYSILKLFKDMKNDXYI
EYNTPLHNACTAGNIDIISIVLYIGCNIPLTKDNKSFIDTISQPDNIYRNCIIYH
LIRNGLDISIVSKLSLIIPLYLTDYVRYIITYCILMNNNFTKYNNOCLNHNHYG
LEFNIJTFDIDNIVSDVNDINRLQKNYNIILRYEDIKYKVLAVNDIYVNTKTFPY
TDIIENCNVMRKKYKILNIVIDKIYSSIVSDNNKLSLLPEIIRIILSKLSEYDLNT
ILYGRDLHKHYKFN"
38873. .39841
/gene="K2R"
38873. .39841
/gene="K2R"
/codon_start=1
/protein_id="CAA75300.1"
/db_xref="GI:3097045"
/db_xref="SPTREMBL:072761"
/translation="MMKWTPSYILLVYVGVSDVPVEHNGKNCGTQDYNNSNNICL
CKQCDPMYTHSCNTTNTKCAKCPDGTFTSIHNHPTCLSCRKGCSSNOQVETKSCS
NTQRECVASGYCEFECSNGCRCLVPQKCDSGYGVYSSKGDVICKKCPGNIDK
CDMSFNSIDVEINMYPVNKTSNCSVSGSSSTISTISELTITLTKHEDCTPVFIGDIYSV
DKLATSGFTNDKVHODLTQCKINLEIKCNSGGESQLTPTAKVYTFMPHSETVTVVG
DCLSNLDVYIVYANTDAIYSDMDVYVHTSYKLVNDHIPNDCCERD"
39919. .40422
/gene="K3R"
39919. .40422
/gene="K3R"
/codon_start=1
/protein_id="CAA75301.1"
/db_xref="GI:3097046"
/db_xref="SPTREMBL:072762"
/translation="MTKVIILGFLIINTNLSMKCQGVSYVNSQELCKCKLCKLGT
YDSHRCDYLLTICGHCPSDTFTSIYNRSPMCBSCGPGCTNRVEVTPCTPTNRRICH
QCKPMTK"
40582. .41214
/gene="S1R"
40582. .41214
/gene="S1R"
/codon_start=1
/protein_id="CAA75302.1"
/db_xref="GI:3097047"
/db_xref="SPTREMBL:072763"
/translation="MATPSLACSIEDDFNNGSVASASVHIRMAFLRKVYGLICLQ
FLTTATTAFLVFLYDFRTIQGSPVLILASMFSGISGLIFALTLLRHKHPLUNYLLCG
FTLSESLTLASVTVFYDVHVVMQAFMLTTAFLALTYYTLQSKRDFSKLAGLFAALW
ILILSGLLRIFVQNETVELVMSAFGALVFCGFIYDTHSLHLKLSKSMC"
41537. .41998
/gene="H1R"
41537. .41998
/gene="H1R"
/codon_start=1
/protein_id="CAA75303.1"
/db_xref="GI:3097048"
/db_xref="SPTREMBL:073582"
/translation="MASPCAQFRPCHCHATKDSLNTVADVHRCHUTEVILWVSHRWTHR
ESAGLYRLIJSITDATELFGSELKDFSDNIDNCVEILKCRINDRSTABE
LRAIIGTQSAIVSGRVFNDKYIDILLMLRKLILNENDYLTLLDHIRTAKY"
42240. .44258
/gene="H2R"
42240. .44258
/gene="H2R"
/codon_start=1
/protein_id="CAA75304.1"
/db_xref="GI:3097049"
/db_xref="SPTREMBL:073579"
/translation="MDYCGRECTILRRLLEDVTTYKKIKLEIETCHNLKSHIDRRGNN
ALHCYVSNKCDTDIKIVRLILSRGVRRLRRNEGLTPIGAYSKRRHVKYQIVHLLTSS
YSNSNELKSNINDFOLYSYMSNDIDLRLIKYIVDKRIPRPSKNTWYANGLGVLVI
YVTPPNRPEVLLWLKSECYSTGYVFRTCMYDSMDCKNSLHYIYILSHRESQSLSKDV

IKCLIDNNVSTHGDEGSLPIQYVWSCSTIDIEIVKLLIKDVDCRVYDDISQPIYR
GVLADYLNKRVRTPYNDVMEIVNLLIEGRHLLDVMKRSITSYDSREKHYIIDNLIK
RFRQDESIVQAMLLINYLHYGDMVSIPIVRCMLDNGATMDKTDNNYPLHDFYVNNILK
VDDNVIRENNHGMVHVSNNRGKCMYGLILSRFNCGYHCYETLIDVDFDILSK
YMDNDIDNENKTKLLYYAVDVNNIQIAKRLLEYGANVNATSLSTINIAIKSSVYRE
NKTLLDULLSVHPTLETMDAFNRDIRLYLPEPLACIRVALILDDDFPSKVYDIS
GRHKLERYADINRKNAYISGVSMEDILFKRSKRHLRYAKNPTFNFVNSNIAWK
KELSIITETVNGERDISVINDNLLISKLPMEIQREILLYYSIK*

gene

/gene="H3R"

CDS

/codon_start=1

/db_xref="GI:3097050"

/db_xref="SPTREMBL:073560"

/translation="MDMEDEIVRVNDMMVIPNAFMDGGENEGHISVNNVCHMYLAF

FDVSISSLFLKLVIKHCDLNLKRGNSPLCHYTMNTRFKPSVLKILLHIGMRNFDK

DEKGIHPLHVLHLSLSDNKKIFDILDPIDDFSKSDLLCYLYRFENGSLNYVLY

KLLKGDSPNCVDEGLTSLHYCKHISTFHESNYKSSHTKMAERKFYIITDHGA

NINAVTIGNTPLHTYLOQYTKSPRVYALLSRGADTRIRNNLDCTPIIMEYIKNDCA

TGHTLIMLLNHEQYKGLQKEGCHLLYLFKHNQGYGSHTLNLYLLDRFDIOKD

EYNTMTPLHTAFQCNNNVASYLVICYDINLPTKDDKTVDFVDFENRNLIYKADVV

NDIITHRLKYSPLMKSLFYKMSFEFSDYVYVKKIAYCLLRDESFAELHSKFLNE

KLSVFWKNTISFDKIDSLIEKSRDISRLKIRISLDDLTIVLTEDIRVITYLEALH

SDRHSFPMWDDLIEQCHLSMEHKSCLVDNALNKESTIGOSRUSYLPPEIMRNLIITK

LSDTHLNMLYGLNHYKYYIP*

/gene="H4R"

/gene="H4R"

/codon_start=1

/protein_id="CAA75305.1"

/db_xref="GI:3097051"

/db_xref="SPTREMBL:073559"

/translation="MKSVLYSYILFLSCLIIINGRDIAHPAGNCKGCKDNEVNRHNLCC

LSCPQTVASRLCDSKTNTOCTPCGSGFTSRNHLPACLSCNGRCDNSQVETRSC

NTNHRICECAPGYCLLKSSGCKACVQSKGIGYGVSGHTSTGDNVCSPLGLGY

SHTVSSADKCEPVPSTFNFDVNLNYPNDVSCRTTITGLSEISTSELITMNH

KCDPFRDGYFVNLKVATVGFPTGENTYISKVCTLNFKEIKCNKNDSSSKQLTKT

KNDGIMPSETVTLVGDCLSVDIYILYSNTNTQDYETDITISYHVGNVLDVDSMPG

SCDIHKLITNSKPTRFL"

/gene="H5R"

/gene="H5R"

/codon_start=1

/protein_id="CAA75307.1"

/db_xref="GI:3097052"

/db_xref="SPTREMBL:073568"

/translation="MKQYIVLACMLAAAMPASLQOSSSCTEENKHHMGIDVITK

VTKQDQPTNDKICQVTEITESDPEVSSDSTSVEDVDPPTTYYSIIIGGLR

MNFQFTKCPQIKSISADGNTVNARLSSVPGQKDSPAITHEEALAMIKDCVSTID

IRCSSEKSDSDIKTHPVLGNSIHHKVSVEDIIGSTIVDTKCVKNLEFSVRIQDMCKE

SSELVKDGPKYVDGSASEGATDDTSLIDSTKLKACV"

BASE COUNT 17414 a 8255 c 8702 g 15278 t

ORIGIN

CVY15035 Length: 49649 February 15, 2002 13:44 Type: N Check: 6849 ..

Found using 'seq4' (pappu403.key)

...

43646 GGTGATGTCAATAATACAAATTGCAAAAGCGGTTATTGGAATATGAGCGACGTTAA

43706 CGCGACGTCGCTCGATATCAATCAATCAGCCATCCAGAAAGTAGTTA

...

10 matches found in sequence:

nma42491 ; TOIG of: nma42491 check: 8990 from: 1 to: 331801
(from "Immun_se.seq")
TOIG of: nma42491 check: 8990 from: 1 to: 331801

LOCUS NMA42491 331801 bp DNA BCT 04-DEC-2000
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 4/7.
ACCESSION AL162755 AL157959
VERSION AL162755.2 GI:7379742
KEYWORDS
SOURCE Neisseria meningitidis Z2491.
ORGANISM Neisseria meningitidis Z2491
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.

REFERENCE 1 (bases 1 to 331801)
AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,I., Davies,R.M., Davis,P., Devlin,K., Felwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
Nature 404 (6777), 502-506 (2000)

JOURNAL 20222556
MEDLINE 2 (bases 1 to 331801)
REFERENCE Direct Submission
AUTHORS Parkhill,J.
TITLE Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
JOURNAL sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENT Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES
source
1. 331801
/organism="Neisseria meningitidis Z2491"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
117..1333
/gene="NMA1040"
117..1333
/gene="NMA1040"
/note="NMA1040", possible type I restriction-modification
system specificity protein, pseudogene, len: 1217 bp;
N-terminus shows weak similarity to the C-terminal half of
TR.Q50359 (EMBL:U25415), hsdS1B, Mycoplasma pulmonis
restriction-modification enzyme subunit S1B (336 aa),
fasta scores; E(): 2.3e-06, 26.2% identity in 141 aa
overlap and to many hypothetical restriction-modification
subunits. Also similar to NMA1041, fasta scores; E():
4.8e-10, 37.0% identity in 100 aa overlap. C-terminus
similar to part of SW:Tris1_ECOLI (EMBL:X13145), hsdS,
Escherichia coli type I restriction enzyme EcoRI24II
specificity protein (410 aa), fasta scores; E(): 1.9e-14,
39.6% identity in 182 aa overlap. Similar to NMA1040, E():
2.9e-09, 38.0% identity in 100 aa overlap. Contains pfam
match to entry PF01420 Methylase_S, Type I restriction
modification DNA specificity domain. Contains a G(8) tract
which would allow translation as an intact CDS, if
variable. Lies within a region of unusually low GC
content"
/codon_start=1
/pseudo
/transl_table=11
/product="pseudogene (putative type I
restriction-modification system specificity protein)"
complement(157..166)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
171..620

gene

CDS

BASE COUNT

ORIGIN

CVY15035 Length: 49649 February 15, 2002 13:44 Type: N Check: 6849 ..

Found using 'seq4' (pappu403.key)

...

43646 GGTGATGTCAATAATACAAATTGCAAAAGCGGTTATTGGAATATGAGCGACGTTAA

43706 CGCGACGTCGCTCGATATCAATCAATCAGCCATCCAGAAAGTAGTTA

...

10 matches found in sequence:

/gene="NMA1040"
/note="Pfam match to entry PF01420 Methylase_S, Type I restriction modification DNA specificity domain, score 105.50, E-value 1.1e-27"
699. 706
/note="(9)8"
726. 735
/gene="NMA1040"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
740. 1231
/gene="NMA1040"
/note="Pfam match to entry PF01420 Methylase_S, Type I restriction modification DNA specificity domain, score 158.70, E-value 1e-43"
1405. 1408
1415. 4509
/gene="NMA1042"
1415. 4509
/gene="NMA1042"
/note="NMA1042, pseudogene, probable type I restriction-modification system restriction protein, len: 291 aa; similar to parts of many e.g. SW:TrlRl_ECOLI (EMBL:X13145), hsdR, Escherichia coli type I restriction enzyme Ecor124II R protein (EC 3.1.21.3) (1033 aa), fasta scores; E(): 0, 73.9% identity in 291 aa overlap (frame 1), followed by a stop codon, E(): 0, 92.9% identity in 127 aa overlap (frame 1), followed by a frameshift near a poly-A tract, E(): 0, 72.5% identity in 604 aa overlap (frame 3)"
/codon_start=1
/pseudo
/transl_table=11
/product="type I restriction-modification system restriction protein (pseudogene)"
complement(1642. .1651)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1687. 1696
/gene="NMA1042"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
1761. 1770
/gene="NMA1042"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
3906. 3915
/gene="NMA1042"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4514. 4545
/note="Stem loop containing DNA uptake sequences: aaat gccgtctgaa at tcagacgac attt"
4519. 4528
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(4531. 4540)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(4559. .6838)
/gene="clpA"
complement(4559. .6838)
/gene="clpA"
/note="NMA1045, clpA, probable ATP-dependent protease ATP-binding protein, len: 759 aa; similar to many e.g. SW:CLPA_ECOLI (EMBL:M31045), clpA, Escherichia coli ATP-dependent Clp protease ATP-binding subunit (758 aa), fasta scores; E(): 0, 56.2% identity in 762 aa overlap, Similar to NMA1683, fasta scores; E(): 0, 37.2% identity in 844 aa overlap. Contains Pfam match to entry PF00495 clpA_B, Chaperonin clpA/B, PS00870 Chaperonins clpA/B signature 1, PS00871 Chaperonins clpA/B signature 2 and two PS00017 ATP/GTP-binding site motif A (P-loop)"

/codon_start=1
/transl_table=11
/product="putative ATP-dependent protease ATP-binding protein"
/protein_id="CAB84311.1"
/db_xref="GI:7379743"
/db_xref="SPTREMBL:O9JV15"
/translation="MLSPELHILQLLYREARNARYEFISLEHLLLVLEEDAAPNV LKLCGADLKLVSQEAASVAENTPQIPDHLDTVETQPTLGFQIRAMVHTQSAGK AAVEPDLILVAMSESESHAVYFLKQSIITRFEVLRCIAHSGSPDDGNDSDGLGREG EEAQKTSLSDYTYNMLNAEVKAGRDLPLGRKHMERLVQILCRKKNPLLVGAG VKTLALAEGLAQHOIVAGDIPDALKDAEVSLLAGTKYRGDFEARKVSKVQLQLE DHIALSRFKQIDVVEPTVSETQILRGLKPMFEGFHQVTRYQCALAAELSAARYNE RFLPKAIDVMEAGAAQRIIPKSKQKVIGRAQIEVTIAKVARIPEKTVSHDKQVL QELGRDLKNMYGOENAIYALVAAVKMSRSGGLGPKDPGIGSFLFSGPTGVGKTEVAKO LAYSGVPLQRPDMSEYMERHAVSRLLIGAPPGYVFEQGLLTEAVNKPCHVCVLLLDE IEKAHPDIFNVLLQVMDAGKLTSDNNGSADFNRNVILMITTNAEASLSRPSIGFTAKR ERGDEMQAINKLFTPEFRNLDAITFPAPLSEPIIVKVVDFLLQLLEHLLDKKVEAE FTPALRKYLAEGFDPQMGARPMRLIQEKIRKPLADELLFGKLVDPGGFVRIWDAAK EEAVLKFKSKVKVPETV"
/gene="clpA"
/note="Pfam match to entry PF00495 clpA_B, Chaperonin clpA/B, score 583.70, E-value 1.1e-171"
complement(5228. .5284)
/gene="clpA"
/note="PS00871 Chaperonins clpA/B signature 2"
complement(5930. .5353)
/gene="clpA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(5894. .5932)
/gene="clpA"
/note="PS00870 Chaperonins clpA/B signature 1"
complement(6173. .6196)
/gene="clpA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
6374. 6383
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(6842. 7153)
/gene="NMA1046"
complement(6842. 7153)
/gene="NMA1046"
/note="NMA1046, len: 103 aa; unknown, similar to bacterial hypothetical proteins e.g. SW:YLJA_ECOLI (EMBL:AE000190), yljA, Escherichia coli hypothetical protein (103 aa), fasta scores; E(): 7.5e-14, 44.6% identity in 83 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1046"
/protein_id="CAB84312.1"
/db_xref="GI:7379744"
/db_xref="SPTREMBL:O9JV14"
/translation="MNHPTDHQSDTLSDINTQPPKRYGVFLINDDYTTMFPVVEIL TEVFMIAQEQAVAVMLLVHHEGKGLCGTTRDIAQTKHQVMERAKTEGHPKLCIYEE V"
complement(6847. .6851)
/gene="NMA1046"
complement(7171. 7180)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
7399. 7404
7411. 7614
/gene="cspA"
7411. 7614
/gene="cspA"
/note="NMA1047, cspA, probable transcriptional regulator, len: 67 aa; similar to many cold-shock proteins e.g. SW:CSPA_ECOLI (EMBL:M30139), cspA, Escherichia coli cold shock protein (69 aa), fasta scores; E(): 1e-15, 64.5% identity in 62 aa overlap. Contains PS00352 'Cold-shock'

```

DNA-binding domain signature"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB84313.1"
/db_xref="GI:7379745"
/db_xref="SPTREMBL:O9JV13"
/translation="MATGIKWFNDAKGFGFIYPDEGGDLFAHFSAINMDGFKTLKE
QKRVSDFTVTPGPKGQAANIQA"
7453..7512
/gene="cspa"
/notes="PS00352 'Cold-shock' DNA-binding domain signature"
7453..7605
/gene="cspa"
/notes="Pfam match to entry PF00313 CSD, 'Cold-shock'
DNA-binding domain, score 107.50, E-value 1.6e-28"
complement(7889..7898)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(7912..9243)
/gene="NMA1048"
complement(7912..9243)
/notes="NMA1048"
/notes="NMA1048, len: 443 aa; unknown, similar to many e.g.
SW:PMB_A_ECOLI (EMBL:X54152), pmbA, Escherichia coli
protein possibly involved in the control of DNA gyrase
(450 aa), fasta scores; E(): 0, 50.4% identity in 425 aa
overlap. Contains Pfam match to entry PF01523 PmbA_TiDd,
Putative modulator of DNA gyrase"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1048"
/protein_id="CAB84314.1"
/db_xref="GI:7379746"
/db_xref="SPTREMBL:O9JV12"
/translation="MLFNHTASSELLDLCRLTLAKATGATAEAADFSELSGQSVSR
LGEIEIOEQKSLDITVVGKRGAGTADPSEQALRDTVRAADIARHTAEDGCA
GLADPELMAQHIGDPDLYHEWDLDTAAYGLAKQCEQAALNEDERLENSEGAAYQTGH
QYVYGYTGFAHRGOSTHHSICSVAADEMGQDIWYDSACRHPDMDSPEITGOT
AARTURRLGSRISPIPTGNYPVLFTTVSGGLHGLVGLSGGALYRQSSFLIDSIGKK
VPDFLNLREEPHPSIFRSSFYDAEGVATAPRFVQNGIVEGYFLSSYSARKLMGT
TGAGAGHLYLNHTHETQSDLLKEMGTGLLVTELMGQAGNTITGDYSRGAAGFEWEN
GVIAYPVHEITVAGRLQDMYRDTGVGVADDAALRSSNKIGSILIAGMTVAGS"
complement(8284..9165)
/gene="NMA1048"
/notes="Pfam match to entry PF01523 PmbA_TiDd, Putative
modulator of DNA gyrase, score 182.80, E-value 5.6e-51"
9334..9909
/gene="NMA1049"
9334..9909
/gene="NMA1049"
/notes="NMA1049, len: 191 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YJGA_HAEIN (EMBL:U32794),
H1151, Haemophilus influenzae hypothetical protein (178
aa), fasta scores; E(): 9.4e-19, 36.4% identity in 176 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1049"
/protein_id="CAB84315.1"
/db_xref="GI:7379747"
/db_xref="SPTREMBL:O9JV11"
/translation="WPSERGGILLCAIGKNRSTKMFQEDSEWIKTMKKQMNGLOD
LMEITKSDNLTLLIGLDELDYEAIVTKITSNGALKRQAOFTGLRMRDTPAPIE
AFELAKRGDAAHNAFLQREVEARVRLDADGALTQFMSDFPHADAGKLTIRLTKK
EQEQNKPKNFALFQELKTVMSQGGTGEA"
9336..9345
/gene="NMA1049"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
9956..10669
/gene="NMA1051"
9956..10669

```

```

/misc_feature
stem_loop
misc_feature
misc_feature
gene
CDS
misc_feature
misc_feature
misc_feature
gene
CDS
misc_feature
misc_feature
misc_feature
gene
CDS
misc_feature
gene
CDS

```

/gene="NMA1051"
 /note="NMA1051, possible membrane protein, len: 237 aa;
 unknown, contains probable N-terminal signal sequence and
 hydrophobic, possible membrane-spanning region"
 /codon_start=1
 /transl_table=11
 /product="putative membrane protein"
 /protein_id="CAB84316.1"
 /db_xref="GI:7379748"
 /db_xref="SPTREMBL:O9JV10"
 /translation="MLFRKTTAAVLAATLMLNCTVMWGMNSPFSFTTARKKHVDKQ
 IRAFGVAEDNAOLEKSGVMGKYWFVNPEDSAKLTLKAGLDKQFQMWEPNPR
 FAYQALPKLESASPQSTEGICLDYRDPADIAKLQLEFEAEVLDRTIYTRCV
 SAKGYATPQKLNADYHFQSPVADIYTVTKKHTDKSLFENIATPTTLILDVAG
 AVIALPVAALIAATNSDDK"
 10673..10707
 /note="Stem loop containing DNA uptake sequences: gcaat
 gccgtctgaa aagcc ttcagacgac attc"
 10678..10687
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 complement(10693..10702)
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 10738..12438
 /gene="recJ"
 10738..12438
 /gene="recJ"
 /EC_number="3.1.-.-"
 /note="NMA1052, recJ, probable
 single-stranded-DNA-specific exonuclease, len: 566 aa;
 similar to many e.g. SW:RECJ_ECOLI (EMBL:M54884), recJ,
 Escherichia coli single-stranded-DNA-specific exonuclease
 (EC 3.1.-.-) (577 aa), fasta scores; E(): 0, 47.9%
 identity in 584 aa overlap. Contains Pfam match to entry
 PF01368 DHH, DHH family"
 /codon_start=1
 /transl_table=11
 /product="putative single-stranded-DNA-specific
 exonuclease"
 /protein_id="CAB84317.1"
 /db_xref="GI:7379749"
 /db_xref="SPTREMBL:O9JV09"
 /translation="MSVKIOTSVNTDFVNHLLTAGADPLIARLCASRGVQSPAEILD
 KLASLLPYQTLNCEAAARLADAVQERKILIVADGATACAGVLDGLAAMGAK
 VDFLVPNREFHGVLTPELAETAAGQVLLITVDNGIASIAGVAAQALGLDIVTD
 HILPAETVPDCITVNPQKCGPSPKSLAGVGVFVLMALRAELRRRNVFSGLEP
 NLGDLDDLVALGTVDVSLDHNRLVYSGQLKMRSGMRGICRIFALFEVARWRKA
 QPFMGFALGPRINAGRLDDMSVGIACLLARDSDSEALAAQLNNIERREIQSM
 LDALNAFPETLPFGQMTLVAYRDDPHOGVGVIVASRLKDRFYRPTIVFAPADNGVR
 GSGRSIPNLHLRDLALDLVSKRHPDLILKFGHAMAAGLSLEHNIPIFQVTFPEAVRE
 MYCEDDLSTQFTITDGSIPACDITLEQAONLARHVWGQGFAPTSFTDFEHVVRQPLGA
 EGKHKKVVLLQKDGCEFEANFWRCSEDIPEYIRTVYRVPVANEWRNNLEQLIYDYWEAA
 "
 10939..12084
 /gene="recJ"
 /note="Pfam match to entry PF01368 DHH, DHH family, score
 351.60, E-value 8.6e-102"
 complement(11370..11379)
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 12686..12695
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 12739..14100
 /gene="pcnB"
 12739..14100
 /gene="pcnB"
 /EC_number="2.7.7.19"
 /note="NMA1053, pcnB, probable poly(A) polymerase, len:
 453 aa; similar to many e.g. SW:PCNB_ECOLI (EMBL:M20574),
 pcnB, Escherichia coli poly(A) polymerase (EC 2.7.7.19)
 (472 aa), fasta scores; E(): 0, 41.3% identity in 441 aa

overlap"
/codon_start=1
/transl_table=11
/product="putative poly(A) polymerase"
/protein_id="CAB84318.1"
/db_xref="GI:7379750"
/db_xref="SPTREMBL:Q9JV08"
/translation="MLKWLNMPLSGRSSKKAESKTVIPAERHNIRAEMLSFAAENV
IRRLKAGQAYVVGGAIRDLILGIEPKDFVATDAMPEQVHKLFRSRIRGRRFOIV
HYMNGAEITEVTFRGGAQVHONAGRIMKNTYGSIEDAMRDFTCNALYDPEKE
EILDHNGJADVAARLVMIGDAERYQEDPVILRAIRLSKGLGFELSEETAAPIAE
SICRLKHEPVARLEDEIMKLLFSGHARECLKRLNGFDIPDDIHPLLNALRVSDGIAGK
MTVLAKNTDERLADKSVSGFVLAALMPELERHWSNLOQGLKPPVALSDAINTM
RETVERGWGPQRFQSATMREIWMFQFQFENRKGARPHKLFQAQRFRAAYDFLLLRKET
GNADRALAEWTAFTQASQEQRTEMTKNEAAAHKEKNEQAKRRRRRRKPKVVGVT
DNE"
12875..12884
/gene="pcnB"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
13400..13409
/gene="pcnB"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/note="Core DNA uptake sequence: gccgtctgaa"
complement(13545..13554)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(13955..13964)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
14125..14277
/note="REP 2: hmms hit to HMM REP 2 (1 - 128), score:
105.80"
14262..14265
14270..14596
/gene="NMA1054"
14270..14596
/gene="NMA1054"
/note="NMA1054, possible periplasmic protein, len: 108 aa;
unknown, contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84319.1"
/db_xref="GI:7379751"
/db_xref="SPTREMBL:Q9JV07"
/translation="MMKKCILGILTACAMPAPAFADRGIDLEARLAQLEHRVAVLESGS
NTVKIDLFGSNTMYCVSTPFPQKTFEASDRNEGVARQVQACNRETSAMFCEDEAI
RCRKKFD"
complement(14682..14691)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(14701..15654)
/gene="NMA1056"
complement(14701..15654)
/gene="NMA1056"
/note="NMA1056, len: 317 aa; unknown, similar to many
bacterial proteins of the phoH family e.g. SM:PHOL_ECOLI
(EMBL:AE000170), ybeZ, Escherichia coli hypothetical
protein (359 aa), fasta scores; E(): 0, 58.1% identity in
320 aa overlap. Contains PS00017 ATP/GTP-binding site
motif A (p-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1056"
/protein_id="CAB84320.1"
/db_xref="GI:7379752"
/db_xref="SPTREMBL:Q9JV06"
/translation="MTHTVHLHLEETDNLALQCGSFONNLDLLAKALDIIHISRRFE
HFTNGAFAGKRALKILLETACTRDLDNGDILRAAVEAOTEDACHOENHDHAYYF
RTRKRSIGGTRPRQNGYIRALLNHDIVFGLGPACTGKTYLAAVAADVANEKHOVERII
LYRPAVEAGEKGLFPGDITQKVPYLRPLYDALYDLMGFDRDKLTIEKGLIEIAPLA
YMRGRTLNGAYITLDEAQNNTPEQMKMFLTRIFGGAKEAVITGDTSQIDILPKNIKISGLK

DAREKLHGVEGLYFHTTGTGEDVVRHPLVQKIVEAYESAHDH"
complement(15235..15255)
/gene="NMA1056"
/note="PS00017 ATP/GTP-binding site motif A (p-loop)"
complement(15694..15849)
/note="Correia element; hmms hit to HMM Correia (1 -
136), score: 272.96"
/label=Correia
complement(15793..16485)
/gene="NMA1057"
complement(15793..16485)
/note="NMA1057, possible glycosyl transferase, len: 230
aa; shows weak similarity to many putative glycosyl
transferases e.g. to the C-terminal half of TR:AAD35706
(EMBL:AE001736), TW0622, Thermotoga maritima putative
lipopolysaccharide biosynthesis protein (388 aa), fasta
scores; E(): 1.5e-15, 31.9% identity in 226 aa overlap and
SW:WCAL_SALTY (EMBL:X56793), wcal, Salmonella typhimurium
putative calanic acid biosynthesis glycosyl transferase
(406 aa), blastp score: Expect = 1.0e-08. Contains Pfam
match to entry PF00534 Glycos_transf_1, Glycosyl
transferases group 1. GC frame and codon usage plots
indicate a change of coding frame at codon 21 (approx.),
near a poly-A tract. This suggests that NMA1058 may
represent the original N-terminal half of this CDS, now
frameshifted. Alternatively, they may both be unrelated
pseudogenes"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase"
/protein_id="CAB84321.1"
/db_xref="GI:7379753"
/translation="MAVPPNIIIGLIFNRPFQREKNKPHFVFCVSHLRLLKGHGV
LITAFARALAQCPQLRLNIGGSGQEORLKQQAADLGITHAVTFGLAQPEAVLDLMR
NSDAFVLASRTETFGVYIEALNSQGLPVIATRCGGAESIVSDGNGLYVPVDDDDALAD
ALIKMYEHSDFEPARLENFGENAVIGRLIGIFRQATAEYKGIKIPVKYSGLTK
IITRRRSRQYK"
complement(15922..16446)
/gene="NMA1057"
/note="Pfam match to entry PF00534 Glycos_transf_1,
Glycosyl transferases group 1, score 90.40, E-value
9.5e-25"
16051..16060
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(16302..16994)
/gene="NMA1058"
complement(16302..16994)
/gene="NMA1058"
/note="NMA1058, len: 230 aa; unknown, shows weak
similarity to part of TR:O32272 (EMBL:Z99122), tuaC,
Bacillus subtilis hypothetical protein (389 aa), fasta
scores; E(): 0.0033, 30.3% identity in 132 aa overlap. GC
frame and codon usage plots indicate a change of coding
frame at codon 189 (approx.), near a poly-A tract. This
suggests that NMA1058 may represent the original
N-terminal half of NMA1057, now frameshifted. TR:O32272
has a Pfam PF00534 Glycos_transf_1, Glycosyl transferases
group 1 domain in its C-terminal half (absent from this
ORF, but present in NMA1057)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1058"
/protein_id="CAB84322.1"
/db_xref="GI:7379754"
/db_xref="SPTREMBL:Q9JV05"
/translation="MWTGFFSKSGTGAIBERHQNRRACDVPLLAERNKSHDPDPYGF
AYRSGDLGIYAHGVIYFFRPFDIDICIRWCAGLKAFKPIRNGNPLDIHAHRM
NYAGILIAOKISQYIGIPYVVTBHSSTITGLIRPHOMQMEKAAHASALLAVSRHFA
HVLQHYGGEWQYLPPTYWAEYSTDLNKKTKTNRISCSAPRFTAVSKDTSYSLP
LGRMHNARNC"
17175..17178
RBS

gene

17186..17890
/gene="NMA1059"
17186..17890
/gene="NMA1059"

CDS

unknown, contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11

/product="putative periplasmic protein"

/protein_id="CAB84323.1"

/db_xref="GI:7379755"

/db_xref="SPTREMBL:O9JV04"

/translation="MLRSILAAASLAVSFPTAAALNVLIVERSESACVAEQDTMSA
RFQVTAERGDKNAVAEFVKFNFTKRSNGSKFTELVSRSAPRYQYINGRIQIG
WEAAEFKVGRRNFDALNRFIADVOADALEYTDFHVSRRERNEVIDOVSKDAVLRFK
ARAEKLAGVIGASGKIVKLNGLHIGSHIAGGGAQAQAKMLRAMPMAASVNEGADSA
PGVEETISVNGTVQF"

17905..17949

stem_loop

/note="Stem loop containing DNA uptake sequences: caaat

gccgtctgaa acccgacgataaggg ttacagcggc attta"

misc_feature

17910..17919

/note="Core DNA uptake sequence: gccgtctgaa"

misc_feature

/label=DUS
complement(17935..17944)

/note="Core DNA uptake sequence: gccgtctgaa"

misc_feature

/label=DUS
complement(17954..18520)

gene

/gene="NMA1060"

CDS

complement(17954..18520)

/gene="NMA1060"

/note="NMA1060, len: 188 aa; unknown, shows weak

similarity to SW:DCD-ECOL (EMBL:M90069), dcd, Escherichia

coli deoxycytidine triphosphate deaminase (EC 3.5.4.13)

(193 aa), fasta scores; E(): 0.00015, 28.5% identity in

151 aa overlap. Similar to many hypothetical deoxycytidine

triphosphate deaminases e.g. TR:O9Z8F1 (EMBL:AE001623),

dcd, Chlamydia pneumoniae putative deoxycytidine

triphosphate deaminase (190 aa), fasta scores; E(): 0,

67.6% identity in 188 aa overlap. Contains pfam match to

entry PF00692 dUTPase, dUTPase"

/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1060"

/protein_id="CAB84324.1"

/db_xref="GI:7379756"

/db_xref="SPTREMBL:O9JRE8"

/translation="MSIKSDKWIRRMSEFEFGMDIPFPNOIKADGKRIISYGTSSYG
YDLCANEFKIFTNINSTIVDPKNEPKNFVTVEDDCCIIPNPSFALARTVEYERIPR
NVLTVCLGRKSTVARGCGLIIVNTPPEPEWEGYVLEFSNTPLPAKIYAGEGVAQVLFF
ESDEICEITYKDRNGKYMGTGVTLFKA"

complement(17960..18430)

/gene="NMA1060"

/note="NMA1060, len: 188 aa; unknown, shows weak

similarity to SW:DCD-ECOL (EMBL:M90069), dcd, Escherichia

coli deoxycytidine triphosphate deaminase (EC 3.5.4.13)

(193 aa), fasta scores; E(): 0.00015, 28.5% identity in

151 aa overlap. Similar to many hypothetical deoxycytidine

triphosphate deaminases e.g. TR:O9Z8F1 (EMBL:AE001623),

dcd, Chlamydia pneumoniae putative deoxycytidine

triphosphate deaminase (190 aa), fasta scores; E(): 0,

67.6% identity in 188 aa overlap. Contains pfam match to

entry PF00692 dUTPase, dUTPase"

/codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="CAB84325.1"

/db_xref="GI:7379757"

/db_xref="SPTREMBL:O9JV03"

/translation="MLKHLAFLPAMMFALPAASAVLTYSQRPQGTGEGVKGDKGKPA
GKGTWRCDGRNVTGSKNGKFDGCGVTVAAANRELIIEFPNSDSTKFRNVMVLSGTFFK
KGLAHRFTVSNGETLFIKCENGMIKEVKLPKKN"

complement(18528..18531)

/gene="NMA1061"

/note="NMA1061, possible periplasmic protein, len: 138 aa;

unknown, contains probable N-terminal signal sequence"

/codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="CAB84326.1"

/db_xref="GI:7379758"

/db_xref="SPTREMBL:O9JVO4"

/translation="MLRSILAAASLAVSFPTAAALNVLIVERSESACVAEQDTMSA
RFQVTAERGDKNAVAEFVKFNFTKRSNGSKFTELVSRSAPRYQYINGRIQIG
WEAAEFKVGRRNFDALNRFIADVOADALEYTDFHVSRRERNEVIDOVSKDAVLRFK
ARAEKLAGVIGASGKIVKLNGLHIGSHIAGGGAQAQAKMLRAMPMAASVNEGADSA
PGVEETISVNGTVQF"

misc_feature

17905..17949

/note="Stem loop containing DNA uptake sequences: caaat

gccgtctgaa acccgacgataaggg ttacagcggc attta"

17910..17919

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS
complement(17935..17944)

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS
complement(17954..18520)

/gene="NMA1060"

complement(17954..18520)

/gene="NMA1060"

/note="NMA1060, len: 188 aa; unknown, shows weak

similarity to SW:DCD-ECOL (EMBL:M90069), dcd, Escherichia

coli deoxycytidine triphosphate deaminase (EC 3.5.4.13)

(193 aa), fasta scores; E(): 0.00015, 28.5% identity in

151 aa overlap. Similar to many hypothetical deoxycytidine

triphosphate deaminases e.g. TR:O9Z8F1 (EMBL:AE001623),

dcd, Chlamydia pneumoniae putative deoxycytidine

triphosphate deaminase (190 aa), fasta scores; E(): 0,

67.6% identity in 188 aa overlap. Contains pfam match to

entry PF00692 dUTPase, dUTPase"

/codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="CAB84327.1"

/db_xref="GI:7379759"

/db_xref="SPTREMBL:O9JV01"

/translation="MKPTALVGRPNVNGKSTLFNRLTRTKDALVHDLPGLTRDRHYGH
GKVGSKPYLIDTGGPEPVVDGILHEMAKQTLQAVDEADVFLVDGRTGLTPQDKI
IADLRQSRPPYLAIVNKGEGGNRAVLAAYELALGDPYVISAAGDGVYLLIEDIF
ETPPEKEEEEAKHPVFAVIGRPNVNGKSTLVNAILGEEERVIAPDMAGTTTDSIHDIF
ERGGKFTTIDTAGVRRRGKVDAAVEKFSYIKAMQAVEANAVLVLDAAQODIADQDA
TTAGFALEAGRALVAVNKGWGISSEERREQVKRDIRKLYFLDFAPHTFSALKERGI
DGLFDSITQAYNAAMIKMPTPKITRVLOSAIERQOQPPRAGLVRPKMRYAHQGGMNPV
IVVHGNSLHAISDSYTRILTQTFKAFNLQGTPLRIQYNNVSENYPENADDPKPKKPLR
RVLSNRLIEKREGKEKRNFKKTKVSKKQFSK"

complement(20875..20898)

/gene="NMA1063"

/note="NMA1063, probable GTP-binding protein, len: 485 aa;

similar to many hypothetical GTP-binding proteins and to

TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential

protein (485 aa), fasta scores; E(): 0, 96.7% identity in

485 aa overlap. Contains two PS00017 ATP/GTP-binding site

motif A (P-loop)"

/codon_start=1

/transl_table=11

/product="putative GTP-binding protein"

/protein_id="CAB84327.1"

/db_xref="GI:7379759"

/db_xref="SPTREMBL:O9JV01"

/translation="MKPTALVGRPNVNGKSTLFNRLTRTKDALVHDLPGLTRDRHYGH
GKVGSKPYLIDTGGPEPVVDGILHEMAKQTLQAVDEADVFLVDGRTGLTPQDKI
IADLRQSRPPYLAIVNKGEGGNRAVLAAYELALGDPYVISAAGDGVYLLIEDIF
ETPPEKEEEEAKHPVFAVIGRPNVNGKSTLVNAILGEEERVIAPDMAGTTTDSIHDIF
ERGGKFTTIDTAGVRRRGKVDAAVEKFSYIKAMQAVEANAVLVLDAAQODIADQDA
TTAGFALEAGRALVAVNKGWGISSEERREQVKRDIRKLYFLDFAPHTFSALKERGI
DGLFDSITQAYNAAMIKMPTPKITRVLOSAIERQOQPPRAGLVRPKMRYAHQGGMNPV
IVVHGNSLHAISDSYTRILTQTFKAFNLQGTPLRIQYNNVSENYPENADDPKPKKPLR
RVLSNRLIEKREGKEKRNFKKTKVSKKQFSK"

complement(20875..20898)

/gene="NMA1063"

/note="NMA1063, probable GTP-binding protein, len: 485 aa;

similar to many hypothetical GTP-binding proteins and to

TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential

protein (485 aa), fasta scores; E(): 0, 96.7% identity in

485 aa overlap. Contains two PS00017 ATP/GTP-binding site

motif A (P-loop)"

/codon_start=1

/transl_table=11

/product="putative GTP-binding protein"

/protein_id="CAB84327.1"

/db_xref="GI:7379759"

/db_xref="SPTREMBL:O9JV01"

/translation="MKPTALVGRPNVNGKSTLFNRLTRTKDALVHDLPGLTRDRHYGH
GKVGSKPYLIDTGGPEPVVDGILHEMAKQTLQAVDEADVFLVDGRTGLTPQDKI
IADLRQSRPPYLAIVNKGEGGNRAVLAAYELALGDPYVISAAGDGVYLLIEDIF
ETPPEKEEEEAKHPVFAVIGRPNVNGKSTLVNAILGEEERVIAPDMAGTTTDSIHDIF
ERGGKFTTIDTAGVRRRGKVDAAVEKFSYIKAMQAVEANAVLVLDAAQODIADQDA
TTAGFALEAGRALVAVNKGWGISSEERREQVKRDIRKLYFLDFAPHTFSALKERGI
DGLFDSITQAYNAAMIKMPTPKITRVLOSAIERQOQPPRAGLVRPKMRYAHQGGMNPV
IVVHGNSLHAISDSYTRILTQTFKAFNLQGTPLRIQYNNVSENYPENADDPKPKKPLR
RVLSNRLIEKREGKEKRNFKKTKVSKKQFSK"

complement(20875..20898)

/gene="NMA1063"

/note="NMA1063, probable GTP-binding protein, len: 485 aa;

similar to many hypothetical GTP-binding proteins and to

TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential

protein (485 aa), fasta scores; E(): 0, 96.7% identity in

485 aa overlap. Contains two PS00017 ATP/GTP-binding site

motif A (P-loop)"

/codon_start=1

/transl_table=11

/product="putative GTP-binding protein"

/protein_id="CAB84327.1"

/db_xref="GI:7379759"

/db_xref="SPTREMBL:O9JV01"

/translation="MKPTALVGRPNVNGKSTLFNRLTRTKDALVHDLPGLTRDRHYGH
GKVGSKPYLIDTGGPEPVVDGILHEMAKQTLQAVDEADVFLVDGRTGLTPQDKI
IADLRQSRPPYLAIVNKGEGGNRAVLAAYELALGDPYVISAAGDGVYLLIEDIF
ETPPEKEEEEAKHPVFAVIGRPNVNGKSTLVNAILGEEERVIAPDMAGTTTDSIHDIF
ERGGKFTTIDTAGVRRRGKVDAAVEKFSYIKAMQAVEANAVLVLDAAQODIADQDA
TTAGFALEAGRALVAVNKGWGISSEERREQVKRDIRKLYFLDFAPHTFSALKERGI
DGLFDSITQAYNAAMIKMPTPKITRVLOSAIERQOQPPRAGLVRPKMRYAHQGGMNPV
IVVHGNSLHAISDSYTRILTQTFKAFNLQGTPLRIQYNNVSENYPENADDPKPKKPLR
RVLSNRLIEKREGKEKRNFKKTKVSKKQFSK"

complement(20875..20898)

/gene="NMA1063"

/note="NMA1063, probable GTP-binding protein, len: 485 aa;

similar to many hypothetical GTP-binding proteins and to

TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential

protein (485 aa), fasta scores; E(): 0, 96.7% identity in

485 aa overlap. Contains two PS00017 ATP/GTP-binding site

motif A (P-loop)"

/codon_start=1

misc_feature

19041..19050

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS
complement(19054..19953)

/gene="rdgC"

complement(19054..19953)

/gene="rdgC"

/note="NMA1062, rdgC, len: 299 aa; unknown, highly similar

to SW:RDGC-NEIGO (EMBL:AF058711), rdgC, Neisseria

gonorrhoeae protein which affects pilin antigenic

variation when mutated (299 aa), fasta scores; E(): 0,

99.3% identity in 299 aa overlap. Also similar to

SW:RDGC-ECOLI (EMBL:X76979), rdgC, Escherichia coli

recombination associated protein (303 aa), fasta scores;

E(): 0, 35.1% identity in 299 aa overlap"

/codon_start=1

/transl_table=11

/product="rdgC protein"

/protein_id="CAB84326.1"

/db_xref="GI:7379758"

/db_xref="SPTREMBL:O9JV02"

/translation="MWFKQISFVPLNKEKLPKPEADVLADKLAEAEFTTCOGILDFSEGG
TAPVSFSPVLVFPADFTLRVALKKEKVLPAQVIRIDILEKVAEIQNNERNVGRKKE
QELKEQITDILLPAPFTRSSRTEAVFNTRHGYLLVNAASAKAENITLKLREALGGLLE
ASLPNTKQSPSSLSMTGWLQHCCEGGLSDSCDLKGTGDIVPVVKYSKODLTADDEV
QHVKNKGVITQGLVWRQIAFILTQDFTLKRQYLDVLQFEAESNGDAAGLAFASQ
ILMAEKSVMLEELVSYLGGWQD"

complement(19960..19963)

complement(19984..21441)

/gene="NMA1063"

complement(19984..21441)

/gene="NMA1063"

/note="NMA1063, probable GTP-binding protein, len: 485 aa;

similar to many hypothetical GTP-binding proteins and to

TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential

protein (485 aa), fasta scores; E(): 0, 96.7% identity in

485 aa overlap. Contains two PS00017 ATP/GTP-binding site

motif A (P-loop)"

/codon_start=1

/transl_table=11

/product="putative GTP-binding protein"

/protein_id="CAB84327.1"

/db_xref="GI:7379759"

/db_xref="SPTREMBL:O9JV01"

/translation="MKPTALVGRPNVNGKSTLFNRLTRTKDALVHDLPGLTRDRHYGH
GKVGSKPYLIDTGGPEPVVDGILHEMAKQTLQAVDEADVFLVDGRTGLTPQDKI
IADLRQSRPPYLAIVNKGEGGNRAVLAAYELALGDPYVISAAGDGVYLLIEDIF
ETPPEKEEEEAKHPVFAVIGRPNVNGKSTLVNAILGEEERVIAPDMAGTTTDSIHDIF
ERGGKFTTIDTAGVRRRGKVDAAVEKFSYIKAMQAVEANAVLVLDAAQODIADQDA
TTAGFALEAGRALVAVNKGWGISSEERREQVKRDIRKLYFLDFAPHTFSALKERGI
DGLFDSITQAYNAAMIKMPTPKITRVLOSAIERQOQPPRAGLVRPKMRYAHQGGMNPV
IVVHGNSLHAISDSYTRILTQTFKAFNLQGTPLRIQYNNVSENYPENADDPKPKKPLR
RVLSNRLIEKREGKEKRNFKKTKVSKKQFSK"

complement(20875..20898)

/gene="NMA1063"

/note="NMA1063, probable GTP-binding protein, len: 485 aa;

similar to many hypothetical GTP-binding proteins and to

TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential

protein (485 aa), fasta scores; E(): 0, 96.7% identity in

485 aa overlap. Contains two PS00017 ATP/GTP-binding site

motif A (P-loop)"

/codon_start=1

/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84328.1"
/db_xref="GI:7379760"
/db_xref="SPTREMBL:Q9JU00"
/translation="MAHLEEQQELDNFKYFWTKTKWLFVAILLAAALGYLGVTVYVON
RAASQNEAAVLIANTVERAKNKAQSEINAEIAKIQQSYPHSISAAQATLMAAATEF
DAQRYDVAEGHLKWLNSQKDSLIIQALAAQRLGVLLQOKKYDAALAAALDTPVEADFA
PILMETKGDVYAAOGKSEQALKNYGQALEKMPQDSVGRVQLVQMKLDSLK"
complement(22225..23520)
/gene="hiss"
complement(22225..23520)
/gene="hiss"
/EC_number="6.1.1.21"
/note="NMA1065, hiss, probable histidyl-tRNA synthetase,
len: 431 aa; similar to many e.g. SW:SYH_ECOLI
(EMBL:M1843), hiss, Escherichia coli histidyl-tRNA
synthetase (EC 6.1.1.21) (423 aa), fasta scores; E(): 0,
53.7% identity in 423 aa overlap. Contains Pfam match to
entry PF00587 tRNA-synt 2b, tRNA synthetases class II (G,
H, P and S) and PS00339 Aminoacyl-transfer RNA synthetases
class-II signature 2"
/codon_start=1
/transl_table=11
/product="putative histidyl-tRNA synthetase"
/protein_id="CAB84329.1"
/db_xref="GI:7379761"
/db_xref="SPTREMBL:Q9JU29"
/translation="MAQKIQSGVGMNDLLPVEQKDKFLTAAFWAQEDTVGRWTRYG
YQIIRPIVEQTGLFVRSIGEDVDVVKEMYTFSDNSLSLSPGEGTASCLRAVVE
HNFLYNQKLWYMPFRERPKGRYQFHQVIEALGFEQPDIAETIAMSADLW
EKLGIREYTLLEINSIGNRERAAHVAALVEYLTRVEDKLEDGSKRKLTPNLRVLD
KNPDLQECNAAPRLVYLGEASQNHVAREFKAMPDLGLGLOYIENPRLVRLGDIYNQTV
FEWTKDKLAQATVCGGGRYDGLIEELGKGPAPSGFAMGIERLLLVLYSEYSGLEVNA
ADPYAHMGOEGADLVQMYIAQRLAQRGNVMQHSIGYSLKAKMKKADNSGARFALIV
ADELANGITVLKDMNGAHDQQTVAADLTNTLQWKNNA"
complement(22234..22238)
/gene="hiss"
complement(22282..23337)
/gene="hiss"
/note="Pfam match to entry PF00587 tRNA-synt_2b, tRNA
synthetases class II (G, H, P and S), score 203.10,
E-value 4.3e-57"
complement(22552..22581)
/gene="hiss"
/note="PS00339 Aminoacyl-transfer RNA synthetases class-II
signature 2"
complement(22895..22904)
/gene="hiss"
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(23614..24270)
/gene="NMA1066"
complement(23614..24270)
/gene="NMA1066"
/note="NMA1066, possible periplasmic protein, len: 218 aa;
shows weak similarity to part of TR-O53037 (EMBL:U68399),
hmcC, Haemophilus influenzae protein implicated in
haemocin production (198 aa), fasta scores; E(): 4.3e-12,
33.6% identity in 146 aa overlap. Also highly similar to
other proteins from Neisseria meningitidis, NMA0173, fasta
scores; E(): 0, 70.7% identity in 164 aa overlap and
NMA0767, fasta scores; E(): 0, 88.3% identity in 128 aa
overlap. Contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84330.1"
/db_xref="GI:7379762"
/db_xref="SPTREMBL:Q9JU28"
/translation="MMKFYVFLACVVSLSYRLNAAMPNDNPVYGGIKVQSWKE
RDFNTVKQDLDFSCGAASVATLLNNFYQGTLEYEVLKLDKREOMRASFDNRRIMP
DLGFEAKGYALSPEQLAQKIPVIVYLYKRYKDDHFSVLRGIDGNTVLLADPPSLGHVSM

gene
CDS
complement(24328..24519)
/gene="NMA1067"
complement(24328..24519)
/gene="NMA1067"
/note="NMA1067, len: 63 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1067"
/protein_id="CAB84331.1"
/db_xref="GI:7379763"
/db_xref="SPTREMBL:Q9JU27"
/translation="MLPTRSVSSLLLGYLEFEENGTKIAAETDSAWHFLGAACRISI
NGKYYAGNRIVWFAKKAS"
complement(24552..25088)
/gene="NMA1068"
complement(24552..25088)
/gene="NMA1068"
/note="NMA1068, len: 178 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1068"
/protein_id="CAB84332.1"
/db_xref="GI:7379764"
/db_xref="SPTREMBL:Q9JU26"
/translation="MEKIIDSKVRETNFNQAMRVIRAMQIDNISDEQVIVRMSCYIK
KDFIKNIERYQSSDIKSPFPLNLPKPYFYLYDTHLNOIYDLAWSCNDF
LMPEIFILDESIKYIDQDKLFKSRETLYSNFKSIKNQCKNYFCGDFDAPYFKSGV
SOVADYNFVPDAFLFFI"
complement(25256..25456)
/gene="NMA1070"
complement(25256..25456)
/gene="NMA1070"
/note="NMA1070, len: 66 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1070"
/protein_id="CAB84333.1"
/db_xref="GI:7379765"
/db_xref="SPTREMBL:Q9JU25"
/translation="MOLICADWTGIGSFMTFEKTSQAQYRDMSEISVRNFWNLRTGA
EYVINGRRVYHNEAEASASLR"
25460..25615
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 271.26"
/label=Correia
complement(25571..25990)
/gene="NMA1071"
complement(25571..25990)
/gene="NMA1071"
/note="NMA1071, len: 139 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1071"
/protein_id="CAB84334.1"
/db_xref="GI:7379766"
/db_xref="SPTREMBL:Q9JU24"
/translation="MSFGYLIATSQPELLTKSRGETFSLIMDKMDLWIYFRCEGNI
YTRKNETESCLTERGCKWLKHIYEFNRGSEFISYVLLKKRESEENFAEIVLSIKRN
KILTVRSGLHFDLRNIYRIEMYSGNLNLNQYGVASP"
complement(25997..26000)
complement(26048..26413)
/gene="NMA1072"
complement(26048..26413)
/gene="NMA1072"
/note="NMA1072, len: 121 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
repeat_unit
gene
CDS
complement(23614..24270)
/gene="NMA1066"
complement(23614..24270)
/gene="NMA1066"
/note="NMA1066, possible periplasmic protein, len: 218 aa;
shows weak similarity to part of TR-O53037 (EMBL:U68399),
hmcC, Haemophilus influenzae protein implicated in
haemocin production (198 aa), fasta scores; E(): 4.3e-12,
33.6% identity in 146 aa overlap. Also highly similar to
other proteins from Neisseria meningitidis, NMA0173, fasta
scores; E(): 0, 70.7% identity in 164 aa overlap and
NMA0767, fasta scores; E(): 0, 88.3% identity in 128 aa
overlap. Contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84330.1"
/db_xref="GI:7379762"
/db_xref="SPTREMBL:Q9JU28"
/translation="MMKFYVFLACVVSLSYRLNAAMPNDNPVYGGIKVQSWKE
RDFNTVKQDLDFSCGAASVATLLNNFYQGTLEYEVLKLDKREOMRASFDNRRIMP
DLGFEAKGYALSPEQLAQKIPVIVYLYKRYKDDHFSVLRGIDGNTVLLADPPSLGHVSM

```

/transl_table=11
/product="hypothetical protein NMA1072"
/protein_id="CAB84335.1"
/db_xref="GI:7379767"
/db_xref="SPTREMBL:Q9JUJ3"
/translation="MEKPLTPPFELFEDVSLDFQGLSELEKKIEPQDLMDVYRAF
DSVGNILNFRIVEKQKGFVSTKIKTVFDSADMSDDLFLKCLQSSYKAYFETEP
GLDKRLMKTLIOKCFSC"
complement(26395..26940)
/gene="NMA1073"
/gene="NMA1073"
complement(26395..26940)
/notes="NMA1073, len: 181 aa; unknown, similar to parts
others from Neisseria meningitidis e.g. NMA1083, fasta
scores: E(): 3.3e-17, 65.6% identity in 93 aa overlap,
also to NMA1078. Lies within a region of unusually low GC
content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1073"
/protein_id="CAB84336.1"
/db_xref="GI:7379768"
/db_xref="SPTREMBL:Q9JR36"
/translation="MKKQITAAVMLSMTAPAMANGLDNQAFEDQVHTRADAPMQLA
ELSQKEMKETEGLFLAILGGAIGMWTQHGFSYATTGPASVSDVAIAGGLGAI
GVGAAGKVSFAKYGREIKIGNNMRIAPFNGRTGHPGPPHYHRRVTDNTGKTLPG
GIGHRPWESKSDRSKGNRF"
complement(26946..26950)
/gene="NMA1074"
/gene="NMA1074"
complement(27103..27375)
/notes="NMA1074, possible membrane protein, len: 90 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions. Lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84337.1"
/db_xref="GI:7379769"
/db_xref="SPTREMBL:Q9JRH1"
/translation="MEAVIFFSTLGCILAWIRDPIKSKKILARSLYIIGINVLIS
YVLKILNLSVSDGGGKIVAYILSNLFFWVLVLRSLSKPS"
complement(27383..27388)
/gene="NMA1075"
/gene="NMA1075"
complement(27444..27614)
/notes="NMA1075, len: 56 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1075"
/protein_id="CAB84338.1"
/db_xref="GI:7379770"
/db_xref="SPTREMBL:Q9JUJ2"
/translation="MGLGSTCVCICGRNSGFEGKIGRILINGMRGVAVGTVAGGINGY
AGSTGRKNTDIRR"
complement(27644..27904)
/gene="NMA1076"
/gene="NMA1076"
complement(27644..27904)
/notes="NMA1076, len: 86 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1076"
/protein_id="CAB84339.1"
/db_xref="GI:7379771"
/db_xref="SPTREMBL:Q9JUJ1"
/translation="MFLILTSCLVNOIKETEDYTKTDHRTENTYPLPCNLRTIDPIDQ
TEKQSHSINQRKCYCKVHKNGNFYLVVYFNFSLISEYT"
complement(27997..28151)
repeat_unit

```

```

/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 260.70"
/label="Correia"
complement(28251..28499)
/gene="NMA1077"
complement(28251..28499)
/notes="NMA1077, len: 82 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1077"
/protein_id="CAB84340.1"
/db_xref="GI:7379772"
/db_xref="SPTREMBL:Q9JUJ0"
/translation="MNKNEIVALISKIDGQVSSLENFALSVLREKIIKNTVSSDE
KLIIINDLSGLEFDPEKPKVGETLKIEQIDFLV"
complement(28545..28913)
/gene="NMA1078"
/gene="NMA1078"
complement(28545..28913)
/notes="NMA1078, len: 122 aa; unknown, similar to parts
others from Neisseria meningitidis e.g. NMA1073, fasta
scores: E(): 8.4e-14, 64.6% identity in 79 aa overlap,
also to NMA1083. Lies within a region of unusually low GC
content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1078"
/protein_id="CAB84341.1"
/db_xref="GI:7379773"
/db_xref="SPTREMBL:Q9JUY9"
/translation="MANGLDNQAFEDQVHTRADAPMQLAELSOKEMKETEAGALWFA
PVLGAGGRFALTGTRHGLNQALSRGCVGSNKAIMNTMRNPKITQSQSGRTRTGA
DGVVLSQSKVITTWCKPR"
complement(29055..29300)
/gene="NMA1079"
/gene="NMA1079"
complement(29055..29300)
/notes="NMA1079, len: 81 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1079"
/protein_id="CAB84342.1"
/db_xref="GI:7379774"
/db_xref="SPTREMBL:Q9JUY8"
/translation="MHEKPKYKMKNKYILTKAIGITAISIFVLLNOESSGKTEEPSY
FLMFNLLNLFEEKVTMAAVTALIAAHFIFVALSD"
complement(29357..29512)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 246.57"
/label="Correia"
complement(29529..29645)
/gene="NMA1080"
/gene="NMA1080"
complement(29529..29645)
/notes="NMA1080, len: 38 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1080"
/protein_id="CAB84343.1"
/db_xref="GI:7379775"
/db_xref="SPTREMBL:Q9JUY7"
/translation="MELKVCNREAKASVVTFTYFILLISKKKEIQDYTCQ"
complement(29704..29991)
/gene="NMA1081"
/gene="NMA1081"
complement(29704..29991)
/notes="NMA1081, possible membrane protein, len: 95 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions, lies within a region of unusually low GC content"

```



```

/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84344.1"
/db_xref="GI:7379776"
/db_xref="SPTREMBL:Q9JUY6"
/translation="MDDLIAFLGLSGKYSKISRLSISTAGSIFFAVFFIHILI
IFIVDGNKENLWNFITGIMKDLFLSCVCLLFYLLFAETIRKTINK"
30052..30114
/notes="Correia element; hmms hit to HMM Correia (1 - 62),
score: 78.71"
/label=Correia
/complement(30298..30669)
/genes="NMA1082"
/complement(30298..30669)
/genes="NMA1082"
/notes="NMA1082, len: 123 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1082"
/protein_id="CAB84345.1"
/db_xref="GI:7379777"
/db_xref="SPTREMBL:Q9JUY5"
/translation="MRFYIDNKECDYIFPRKAVNISDDFIQSGFTPEGLITKNSNPX
LFNQYSKMYSKNKQCYEMVTLSDMKNIELLSNDDAYIEFAYSDDFYVLPEIEIDK
RTLKSLNFYIDFGVKYIINK"
30746..30787
/complement(30803..31246)
/genes="NMA1083"
/complement(30803..31246)
/genes="NMA1083"
/notes="NMA1083, possible periplasmic protein, len: 147 aa;
unknown, similar to parts of other proteins from Neisseria
meningitidis e.g. NMA1073, fasta scores; E(): 1.4e-17,
65.6% identity in 93 aa overlap, also to NMA1078. Contains
probable N-terminal signal sequence. Lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84346.1"
/db_xref="GI:7379778"
/db_xref="SPTREMBL:Q9JUY4"
/translation="MKKQITSVMLSMTAPAMANGLDNOTENOVFHTQADAPMOLA
ELSQEMKTEGAAPLVAIGILHAGRFLAQRWTCORVAQAALSRGANVYARTSQAR
AVANQAWERQNVIRHCQKEIHSNYSHFQNSOKIRGHAFYGNKHR"
complement(31252..31256)
/complement(31289..32158)
/genes="NMA1084"
/complement(31289..32158)
/genes="NMA1084"
/notes="NMA1084, possible periplasmic protein, len: 289 aa;
shows weak similarity to TR:O53036 (EMBL:U68399), hmcD,
Haemophilus influenzae protein at haemocin immunity locus
(297 aa), fasta scores; E(): 3.7e-07, 24.6% identity in
289 aa overlap. Similar to part of NMA0171, fasta scores;
E(): 1.2e-09, 54.7% identity in 53 aa overlap. Contains
probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84347.1"
/db_xref="GI:7379779"
/db_xref="SPTREMBL:Q9JUY3"
/translation="MKRIFLPALPLISAYADLPLTIEDIMTDGKWKLETSLYL
NSGNRAELAAPVYIQGTAFSFIPTIEQSGNSNTDMLVGLTGLRYGLTGNTDIYGS
GSLVHEERKLDGNGKTRNKRSDVSLGTSHTFLKDDKNPALISFLESTVYEKSRNKA
SGKSKWLGATVYKADIPVLSLTAARYNGSKTLLSNTKYKAGNYWMLNPNISFAAN
DRISLRTGGTQNLGKQPDRLDGKESARNSTYAHFAGFGFTYTTALNASARNVSGQ
SSSELKFGVOHT"
complement(32163..32167)
32281..32290
misc_feature

```

```

/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
32322..33245
/genes="NMA1085"
32322..33245
/genes="NMA1085"
/notes="NMA1085, len: 307 aa; similar to TR:O07166
(EMBL:296070) Mycobacterium tuberculosis hypothetical
protein (305 aa), fasta scores; E(): 0, 42.9% identity in
282 aa overlap and shows weak similarity to many,
including pseudouridine synthases. Contains Pfam match to
entry PF00849 YABO, Hypothetical yabo/ycec/sfHB family"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1085"
/protein_id="CAB84348.1"
/db_xref="GI:7379780"
/db_xref="SPTREMBL:Q9JUY2"
/translation="MKRRNNPLLVGKPSYLVLPHEKQFYGLPLHLICIRFPFVG
ADDRRLNSGFVVGSDGAALDEHSLFEPGKVMFYRETSESEPRIPFEKILHIDE
HLIVDKPHFLPVPISGSRFLRETLTLRLRLPELQHLNVEDITPLHRLDXTAGVMLL
SHNPATRGAYQTMFONKTYVMKTYEALAPTRTDLPPYLDVVSRLVRGKFFTTQEAEGE
PNAHTTVLEIENRGESLYELTPTGCKKHQLRVHMMGLMPLNDALYPVPVPSAGSED
YRPLALAKKIAFADPLSGRERVFCGCL"
32613..33095
/genes="NMA1085"
/notes="NMA1085"
/notes="Pfam match to entry PF00849 YABO, Hypothetical
yabo/ycec/sfHB family, score 131.20, E-value 1.9e-35"
33125..33134
/genes="NMA1085"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
33259..33293
33306..33315
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(33323..33964)
/genes="NMA1086"
/complement(33323..33964)
/genes="NMA1086"
/notes="NMA1086, len: 213 aa; shows weak similarity to
TR:Q9ZD73 (EMBL:AJ235271), RP471, Rickettsia prowazekii
hypothetical protein (238 aa), fasta scores; E(): 1.4e-10,
30.3% identity in 198 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1086"
/protein_id="CAB84349.1"
/db_xref="GI:7379781"
/db_xref="SPTREMBL:Q9JUY1"
/translation="MLKPETIHISGPAGILETHIPSEQVPARGVAVINHPNPLQGGT
NTNKVIOATAKALSKLGFHCYLPNLRGVSGEGTHDYGRGETODCLAVIDYARAOHPE
APFALSGFSFGYVATFAQARIPDLLLLIGAAVCHYTCRPEPSVNPVAKTLMIHG
AEDEVVEIGALKWAEPQDLPVITAGSTHFFHGKLLIVLRDITLIRFAPVCLNG"
complement(33973..33977)
/complement(34019..34810)
/genes="NMA1087"
/complement(34019..34810)
/genes="NMA1087"
/notes="NMA1087, possible transferase, len: 263 aa; shows
weak similarity to many spermidine synthases e.g.
TR:BAAB1738 (EMBL:AB001598), spe-3, Neurospora crassa
spermidine synthase (291 aa), fasta scores; E(): 6.8e-05,
23.0% identity in 257 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transferase"
/protein_id="CAB84350.1"
/db_xref="GI:7379782"
/db_xref="SPTREMBL:Q9JOL8"
/translation="MAHPYRRLRPAKSGPPEVGISEGNIIRSLHLCSDTVQSSMULD
HPSELVLSYSRAMGWLFTDALPQHITQILGGSGFARMIDTLPDTRQTAYDINPQ
VIAIARNLFELPFEKEGFEIADGAEYIKVFRHNTDVLVDGFDGSEQIIDLAVEEPF

```

```

FRDCRNALSSDGIIVTNWWSGDKRYORFIERLLSVFEGRVLELPAESHGNVAVMAFQS
34295..34304
/label=DUS
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
34976..35051
/note="tRNA Thr anticodon TGT", Cove score 86.45"
35161..35170
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
35193..35984
/gene="panB"
35193..35984
/gene="panB"
/EC_number="2.1.2.11"
/note="NMA1088, panB, probable 3-methyl-2-oxobutanoate
hydroxymethyltransferase, len: 263 aa; similar to many
e.g. SW:PANB_ECOLI (EMBL:L17086), panB, Escherichia coli
3-methyl-2-oxobutanoate hydroxymethyltransferase (EC
2.1.2.11) (264 aa), fasta scores; E(): 0, 49.2% identity
in 262 aa overlap"
/codon_start=1
/transl_table=11
/product="putative 3-methyl-2-oxobutanoate
hydroxymethyltransferase"
/protein_id="CAB84351.1"
/db_xref="GI:7379783"
/db_xref="SPTREMBL:Q9JUY0"
/translation="MITVNTLQMKAAAGKIAMLTAYESSFAALMDADGVLLVGDSD
LCMAVGHQSTLPVSLRDMCYHTECVARGAKNAMIVSDLPFCAYQOQKQEAFAAAEL
MAAGAHMQLGEGVMAETTELQMRGIPVCAHIGLTPQSVPFAGGYKVGQGGRAQ
LLNDKAHDDAGAAVYVLECVPAELAKKVTESVCTPTGIGAGVDCDGGVLLHMDMLG
IFPGATKRVKFMQGSISQIAAVRAYVAEVKAKTFPAAEHIFAD"
35998..36033
/note="Stem loop containing DNA uptake sequences: ccgat
gccgtctgaa agccgt ttcagacgcg atttt"
36003..36012
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
36095..36098
/note="NMA1089, panC, probable pantoate--beta-alanine
ligase, len: 278 aa; similar to many e.g. SW:PANB_ECOLI
(EMBL:L17086), panB, Escherichia coli
pantoate--beta-alanine ligase (EC 6.3.2.1) (283 aa), fasta
scores; E(): 0, 52.1% identity in 282 aa overlap"
/codon_start=1
/transl_table=11
/product="putative pantoate--beta-alanine ligase"
/protein_id="CAB84352.1"
/db_xref="GI:7379784"
/db_xref="SWISS-PROT:P57035"
/translation="MOIITHIQELRAWKNAQTVAFVPTMGNLHGHLLALVREAKKRA
DNVVSIFVNLQFGGDEDDKYPRTLQODADKLEAGVAVVPFAPDEKELYPNVQRY
NVEPNLNLGCKFRGHFLGVATVYSKLFNIVADPVACFGKDKYQQLAVIKGTED
LNFDEIVPDTGRAEDGLSSRNQYLSAAERDEAPRLYRELKAVAESLAQGSIDYA
GLEKRAVQSLTEYGWVVDYVEIRRVDTLEVARAGDKLVLLAAARLGLTTRLIDNLEIK
LP"
36957..36998
/note="Stem loop containing DNA uptake sequences: ggaat
gccgtctgaa gcggatttcgt ttcagacgcg attta"
36982..36971
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
36977..40522
/complement(36984..36993)
/note="Core DNA uptake sequence: gccgtctgaa"

```

```

/label=DUS
37067..37076
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
37099..37104
37108..38946
/gene="NMA1090"
37108..38946
/gene="NMA1090"
/note="NMA1090, possible periplasmic protein, len: 612 aa;
shows weak similarity to hypothetical proteins e.g.
SW:IHE3_PSEAE (EMBL:X82071) Pseudomonas aeruginosa
hypothetical protein (576 aa), fasta scores; E(): 1.7e-15,
26.4% identity in 613 aa overlap. Contains probable
N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84353.1"
/db_xref="GI:7379785"
/db_xref="SPTREMBL:Q9JUX9"
/translation="MLPARETILSVLAAALLAGQAVAGAADAKPKPEKVKPKQQR
YSEEEIKNERARLAAYGERVNOIFTLLGGETALQKQAGTALATYMLLERTKSPPEVA
ERALEMAVSLNAFEQAEEMIQKWRQIEPIPKAKQKAGWLRNVLRENGNHLGDEEV
LQAODEGNRRVFLLLAQAQVODGLAQKASKAVRRAALRYEHLPEAAVADVVFVQGM
REKEKAIQALQRLAKLDTEILPPTLMTLRTARKYPEILDPFPEQDTQNLAVWQGM
EIMNLVSLHRLDDAYARLVNLLERNPNADLYIOAAIILAAKRKESVIDGYAEKAYGR
GTGEQGRGAAATAAMLYADRDYTKVROMLKVKSAPLEYLFDKGYLAAAVALDGGRA
ALQIORVRKLEPQQQRYTADNLKIQFALPKPDKREALGLDKIIEKPPAGSTN
ELQAEQVRSVVYDRIGKKRMSIDLERAFRLAPDQNAQIMNNGYLSLSDSKRLDGG
FALLQTAYQINPDPTANDSIGWAYYLKGDASALPYRLRYSFENDEPEPEVAHLGLEYL
WALGERDQADVMTQAAHLTGDKKIWRETLKRHGIALPOPSPKPRK"
38642..38651
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
38961..38970
/complement(38961..38970)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
39075..39078
39086..39667
/gene="NMA1091"
39086..39667
/gene="NMA1091"
/note="NMA1091, possible outer membrane lipoprotein, len:
193 aa; shows weak similarity to bacterial outer membrane
lipoproteins e.g. SW:LOLB_PSEAE (EMBL:X82071), lolB,
Pseudomonas aeruginosa outer membrane lipoprotein, fasta
scores; E(): 3.5e-05, 26.9% identity in 197 aa overlap.
Contains probable N-terminal signal sequence and
appropriately positioned PS00013 prokaryotic membrane
lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative outer membrane lipoprotein"
/protein_id="CAB84354.1"
/db_xref="GI:7379786"
/db_xref="SWISS-PROT:P57023"
/translation="MKHTVSASVILLLTACAOILPONNENLWQPSHRSFTAEGRLLAV
KAEGKGSVANFDWYQPPVETININIPGLSTIGLCQDQDGLADVGKNGVQAESAE
ELNSRLQVFKLPIDQLHWDAGRVPAGAPYRIPLDQGLEIQYGTWGTADSGGQVRTL
QLNNGLNIRLVFTETIGMPSETETOQCAARTQ"
39101..39133
/gene="NMA1091"
/note="PS00013 prokaryotic membrane lipoprotein lipid
attachment site"
39619..39628
/gene="NMA1091"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
39677..40522
/gene="NMA1092"
39677..40522
/gene="NMA1092"

```

/note="NMA1092, len: 281 aa; similar to many hypothetical proteins e.g. SW:YCHB_ECOLI (EMBL:M77237), ychB, Escherichia coli hypothetical protein (283 aa), fasta scores: E(): 0, 42.4% identity in 278 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1092"
/protein_id="CAB84355.1"
/db_xref="GI:7379787"
/db_xref="SPTREMBL:Q9JUX8"
/translation="MSVTDGROAFPAPAKNLNLDLRITGRREDGYHNIESIFCLIDLQD TVYLPDRDGKILLNPVDGMFQEAADLSYRAASLLQKARTPTGVEIWLDDKPIPTGAG LGGSSDAATVLLNPNWQCGLTQRLDLSGAALGADVFFIFGKNAPFARGIDRLD ENDIPKQWVIVKPVHVSTAKIFTHEGLTRNSASSIMPTFQNLQPFRRNDMAVVFK EYEVKWKAYSELRYGFALMTGSGACVFTACQDRNSAYNIYRQVSDLYEAYLAEGLSKH PLLSV"
40530. 40605
/note="tRNA Gln anticodon TTG, Cove score 71.63"
40618. 40693
/note="tRNA Gln anticodon TTG, Cove score 71.63"
40702. 40777
/note="tRNA Gln anticodon TTG, Cove score 71.63"
40777. 40814
40944. 41927
/gene="prsa"
40944. 41927
/gene="prsa"
/EC_number="2.7.6.1"
/note="NMA1093, prsa, probable ribose-phosphate pyrophosphokinase, len: 327 aa; similar to many e.g. SW:KPRS_ECOLI (EMBL:M13174), prsa, Escherichia coli ribose-phosphate pyrophosphokinase (EC 2.7.6.1) (314 aa), fasta scores: E(): 0, 66.5% identity in 310 aa overlap. Contains Pfam match to entry PF00492 PRPP_synt, phosphoribosyl pyrophosphate synthetase, PS00103 Purine/pyrimidine phosphoribosyl transferases signature and PS00114 Phosphoribosyl pyrophosphate synthetase signature"
/codon_start=1
/transl_table=11
/product="putative ribose-phosphate pyrophosphokinase"
/protein_id="CAB84356.1"
/db_xref="GI:7379788"
/db_xref="SPTREMBL:Q9JQV4"
/translation="NAAYDSLWFTGNAPPELAQRVVRHLDLSLGNASVKESDGEVA VELLENRGRDVFILQPTCAPNDNLMEILFWADALKRASGRITAITPIFYGARQDR RPRSVRPLSAKLVANMLYSAGIDRVLTVDLHADQIQGFDPIDVNIYATPILLNDIK QQRLENLTVPSPDGGVVRARAVAKSLNADLAIIDRRPKANVAEVMNIGDIQGRTC LIVDDMIIDTANTLCKAAVALKERGAERVLAYASHAVFSGEAVSRASSIEDQWVYTDI IPLSEAANKCDRIQRQVTIAGLAEVRRISNEESVSYLNFNEEVMVTGSMLLP"
40962. 41888
/gene="prsa"
/note="Pfam match to entry PF00492 PRPP_synt, Phosphoribosyl pyrophosphate synthetase, score 621.00, E-value 6.8e-183"
complement(41057. 41066)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
41337. 41384
/gene="prsa"
/note="PS00114 Phosphoribosyl pyrophosphate synthetase signature"
41595. 41633
/gene="prsa"
/note="PS00103 Purine/pyrimidine phosphoribosyl transferases signature"
41982. 41985
41994. 42566
/gene="NMA1094"
41994. 42566
/gene="NMA1094"
/note="NMA1094, len: 190 aa; shows weak similarity to SW:CTC_BACSU (EMBL:D26185), etc, Bacillus subtilis general

stress protein (203 aa), fasta scores: E(): 6.7e-11, 30.2% identity in 169 aa overlap and at the N-terminus, to SW:R125_ECOLI (EMBL:D13326), rpiV, Escherichia coli ribosomal protein L25 (94 aa), fasta scores: E(): 3.5e-10, 40.0% identity in 90 aa overlap. Also similar to many hypothetical proteins. Contains Pfam match to entry PF01386 Ribosomal_L25p, Ribosomal L25p family"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1094"
/protein_id="CAB84357.1"
/db_xref="GI:7379789"
/db_xref="SPTREMBL:Q9JUX7"
/translation="MTYEIQASVREAQGTGASRRLRREGQIPGILYGEQGEPPVAIVD HKTVPFALKESFHTALIKSLNGETKDVIVDFQMHFPERREVQHIDFOAVRADQPV RIVPLHIVNAESOAVKLOGRVSLLNTTVEVVALPANIPAFILDCAEVEVAGDILHL SDIKLPEGVSVSLKRNEENLAVATVTGKKR"
42006. 42269
/gene="NMA1094"
/note="Pfam match to entry PF01386 Ribosomal_L25p, Ribosomal L25p family, score 128.60, E-value 1.1e-34"
42617. 42652
/note="Stem loop containing DNA uptake sequences: ttatat gccgtctgaa ccgtgt ttcagacggc atttc"
42622. 42631
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(42638. 42647)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(42691. 43860)
/gene="NMA1095"
complement(42691. 43860)
/gene="NMA1095"
/EC_number="3.4.16.4"
/note="NMA1095, probable D-alanyl-D-alanine carboxypeptidase, len: 389 aa; similar to many e.g. SW:DACC_ECOLI (EMBL:X06480), dacC, Escherichia coli D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) (400 aa), fasta scores: E(): 0, 36.3% identity in 364 aa overlap. Similar to NMA0962, fasta scores: E(): 8.4e-14, 29.6% identity in 243 aa overlap. Contains Pfam match to entry PF00768 Peptidase_S11, D-alanyl-D-alanine carboxypeptidase. Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative D-alanyl-D-alanine carboxypeptidase"
/protein_id="CAB84358.1"
/db_xref="GI:7379790"
/db_xref="SPTREMBL:Q9JUX6"
/translation="MTAHLKILPVLISILGVSHATAASPAPNRPHTAHAAPTFOTPETL TAAHIVIDQSKOILSAKNINTPVEPAALTQMLTAYLVFKNMKSGNIRSEENLKIPES AWASGSRMFVRPGDVTDKLKGMIALSAANDALTLAGRLGNSIENFVQOMKNKE RRLGMKNTVFKNPTGLSREGQVSTAKDLQALSEALMRDPPEYLPFSIKSFKEKNEQ NNRNILLYRDNNVNGVLKAGTSGGYNLAVSYSGNRRHILVITPLGSEATRASDNK LLNWLQAEDTPKVIYPKGTVAQIOISGSGKKTVRAGELKAVITIPHKEAKNAEQIL ETIQIPAPVKKQKILGKIKIKQNGVTIAEKEIVALENVKRSRWORLWACLTGQ"
complement(43000. 43761)
/gene="NMA1095"
/note="Pfam match to entry PF00768 Peptidase_S11, D-alanyl-D-alanine carboxypeptidase, score 233.80, E-value 2.4e-66"
44008. 44534
/gene="ilva"
44008. 44534
/gene="ilva"
/EC_number="4.2.1.16"
/note="NMA1096, ilva, probable threonine dehydratase biosynthetic, len: 508 aa; similar to many e.g. SW:TIDL_ECOLI (EMBL:X04890), ilva, Escherichia coli threonine dehydratase biosynthetic (EC 4.2.1.16) (514 aa), fasta scores: E(): 0, 55.8% identity in 502 aa overlap.

Contains pfam match to entry PF00291 S_T_dehydratase, Pyridoxal-phosphate dependent enzymes, two pfam matches to entry PF00585 Thr_dehydrat_C, C-terminal domain of threonine dehydratase and PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"

/codon_start=1
/transl_table=11
/product="putative threonine dehydratase biosynthetic"
/protein_id="CAB84359.1"
/db_xref="GI:7379791"
/db_xref="SPTREMBL:Q9JUX5"
/translation="MNTPLPSYVNLKILRTASVYDVAVETPLPARSLVRLKNNILL
KRDLOPVFEFKIRGAYNKMSPKLDALACGVIAASAGNHAQGVALSQAQRLGCGRAVIV
MPETPKIKVDVAVKSGGEVVLVGVSYNDAYDYMELAEQEGLYTAPDDPDVIAQG
GTVMELVSHPDPIRAVFPVIGGGLAGVAAFIKQVRPEIKVIGVQTNDCMKQKQ
VEAGEIVHLKDVGLFSDGTAVKYVGNETFRICKLDELDEIITVDVAVCGAVKDFDFT
RSITEPAGALALAGKAYTAREGAQNTLIATVSGAMNPFHRLRHVSESELGEGNEG
IFAVTIPEERSGSLKFVNLILGNRNITFENRYGDEKAHIFVGLQAAGPDOLAVIGSR
LDAGLNPVLDLTDDEIAKHIRYMGVGRDTPKVENRLVSPFEPERPCALARFNLHQG
GWNITLPHYRHGADYGRILVGLDVPDHAADFGLSLGYSYHEETQNAATKFLA
"

44062. .44973
/gene="ilvA"
/note="Pfam match to entry PF00291 S_T dehydratase, Pyridoxal-phosphate dependent enzymes, score 423.10, E-value 2.5e-123"
44146. .44187
/gene="ilvA"
/note="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"
complement(44706. .44715)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
44977. .45246
/gene="ilvA"
/note="Pfam match to entry PF00585 Thr_dehydrat_C, C-terminal domain of threonine dehydratase, score 119.30, E-value 7e-32"
45259. .45528
/gene="ilvA"
/note="Pfam match to entry PF00585 Thr_dehydrat_C, C-terminal domain of Threonine dehydratase, score 131.20, E-value 1.8e-35"
45546. .45580
/note="Stem loop containing DNA uptake sequences: acaat gccgctctgaa agcct ttcagacgac attgc"
45551. .45560
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(45566. .45575)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(45591. .46664)
/gene="cysA"
/gene="cysA"
/gene="cysA"
/note="NMA1097, cysA, probable sulphate permease ATP-binding protein, len: 357 aa; similar to many e.g. SW:CYSA_ECOLI (EMBL:M32101), cysA, Escherichia coli sulfate transport ATP-binding protein (365 aa), fasta scores: E(): 0. 49.4% identity in 360 aa overlap. Contains pfam match to entry PF00005 ABC_tran, ABC transporter, PS00211 ABC transporters family signature and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative sulphate permease ATP-binding protein"
/protein_id="CAB84360.1"
/db_xref="GI:7379792"
/db_xref="SPTREMBL:Q9JUX4"
/translation="MSITQNLNKHFGNFHALKTNLNVPVTKGLVSLGPGCGKTTLL
LRITAGLENADGGKILFDGQDVTKAKVRKRVGVFOHYALFRHMNVFDNVAFGTLVL
PKSPERPSGQIRAKVEELKLVQLSHLAKSYPHQLSGGQQRQRIARALAVEPKLLILL

DEPFGALDAKVRKELRTWLRLDIHNLGVTISILVTHDQEEALEVSEIVVMNHGKIEQT
GSAEAIYRKPENAFVTEFLGETAFEGRIEKGFWHNGFAWKLDQAKWQDOTATGYI
RPHEWQIAAEHETPMICAEIEKHAVGALTHILVKHGKQDVHITLAGSDAARYPIAEG
KELNLIPKQVYVFSONELEIYSI"
complement(46026. .46583)
/gene="cysA"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 234.20, E-value 1.9e-66"
complement(46212. .46256)
/gene="cysA"
/note="PS00211 ABC transporters family signature"
complement(46539. .46562)
/gene="cysA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(46661. .47521)
/gene="cysW"
complement(46661. .47521)
/gene="cysW"
/note="NMA1098, cysW, probable sulphate permease inner membrane protein, len: 286 aa; similar to many e.g. SW:CYSW_ECOLI (EMBL:M32101), cysW, Escherichia coli sulfate transport system permease protein (291 aa), fasta scores: E(): 0. 51.7% identity in 265 aa overlap. Contains pfam match to entry PF00528 BPD_transp, Binding-protein-dependent transport systems inner membrane component and PS00402 Binding-protein-dependent transport systems inner membrane comp sign. Contains hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative sulphate permease inner membrane protein"
/protein_id="CAB84361.1"
/db_xref="GI:7379793"
/db_xref="SPTREMBL:Q9JUX3"
/translation="MKPYSANPNLTPERLRLLVLLIAALGFLLLMLVPLVAVFYBAL
KGGNDIYLSLNDPEAWSAIKLTILITAVIPVNAVIGVAMWLLFRDFRGKQLLT
LIDLPEFSPVAVAGLMPVILLFGAHTGLGMLEAQIQIIFAIPIGLATILFTVFPFVA
REIIPMQAGDSEEQALILGASGWMFWRVTLPNIKWALLYGIILTNARAMEFGA
VSVSGHIRGETNTVPLLVLEIFYNEYNFTGAFALSGLALLATLAVQNIITKLQDK
KLAARNAV"
complement(46672. .46675)
/gene="cysW"
complement(46787. .47026)
/gene="cysW"
/note="Pfam match to entry PF00528 BPD_transp, Binding-protein-dependent transport systems inner membrane component, score 39.50, E-value 7.8e-08"
complement(46937. .47023)
/gene="cysW"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp sign."
47556. .47598
/note="Stem loop containing DNA uptake sequences: gaag gtcgtctgaa aaggattgagggt ttcagacgac cttt"
complement(47586. .47774)
/gene="NMA1099"
complement(47586. .47774)
/note="NMA1099"
/note="NMA1099, len: 62 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1099"
/protein_id="CAB84362.1"
/db_xref="GI:7379794"
/db_xref="SPTREMBL:Q9JUX2"
/translation="MTAARLFYSEFSIKDKRSSENEAPQNTFVGRILVSDKHLRFS
NLVGYKPTYIYLMNKKVV"
47773. .47792
/note=">= 90% match to ATCCCNNNNNNGGAAT"
/label=QRS3
complement(47793. .47850)
/note="RS99"

repeat_unit

/label=RS99
47851..47870
/note=">= 90% match to ATTCCNNNNNNNGGAAT"
/label=dRS3
complement(47899..48735)
/gene="cysU"
CDS
complement(47899..48735)
/gene="cysU"

/note="NMA1100, cysU, probable sulphate permease inner membrane protein, len: 278 aa; similar to many e.g. SW:CYST_ECOLI (EMBL:M32101), cysU, Escherichia coli sulfate transport system permease protein (277 aa), fasta scores: E(): 0, 48.4% identity in 273 aa overlap. Contains pfam match to entry PF00528 BPD_transp.
Binding-protein-dependent transport systems inner membrane component and PS00402 Binding-protein-dependent transport systems inner membrane comp sign. Contains hydrophobic, probable membrane-spanning regions"

/codon_start=1
/transl_table=11
/product="putative sulphate permease inner membrane protein"

/protein_id="CAB84363.1"
/db_xref="GI:7379795"
/translation="MLAKTPGVLPFGKLSLGLTVLCISLLVLPFAMMAKAAEIGW SGFWNTIPNVNLAAMWLSLRMSFYAMLTNVVFGTLVAMVLVRYEPGKGLANALVDL PFLATVATGALATLYAPNGWIGRFFELGKIATPVGWIALVVSLLPFLVRAVQ PVLELSYEYEAATLGSNRTFRRVLLPEITPALTGAGMFPARAGEYGVIFI AGNIPMWSEIPIITIGKLEQFDVGSAVALFMLLVSVFVLNVMQWALGRKSGA KG"

misc_feature

/note="cysU"
/note="pfam match to entry PF00528 BPD_transp.
Binding-protein-dependent transport systems inner membrane component, score 60.40, E-value 3.9e-14"

misc_feature

complement(48163..48249)
/gene="cysU"
/note="PS00402 Binding-protein-dependent transport systems inner membrane comp sign."

repeat_unit

48877..48939
/note="Correia element; hmms hit to HMM Correia (1 - 62), score: 74.65"

repeat_unit

/label=Correia
48941..48983
/note="Correia element; hmms hit to HMM Correia (114 - 156), score: 51.92"

RBS

gene

49016..49019
49026..49358
/gene="NMA1101"

CDS

49026..49358

/gene="NMA1101"

/note="NMA1101, len: 110 aa; unknown, Glu-rich protein"

/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1101"

/protein_id="CAB84364.1"

/db_xref="GI:7379796"

/db_xref="SPTREMBL:Q9JUX0"

/translation="WOILSFQPDIAERMLEGTEGSVNENAQFVRTDNGYTIWAEHGV AALLAPDPPIGIPCFWVEGAELESLEELCVMERGEFDEVEFDGDDDEWLETAQGGGHH GDACACGH"

gene

CDS

49690..50199

/gene="NMA1102"

/note="NMA1102"

169 aa; similar to bacterial hypothetical proteins e.g. SW:Y318_HAEIN (EMBL:U22717), HI0318, Haemophilus influenzae (172 aa), fasta scores: E(): 0, 56.2% identity in 169 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop). Contains hydrophobic, possible

membrane-spanning regions. Lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84365.1"
/db_xref="GI:7379797"

/db_xref="SPTREMBL:Q9JRI2"

/translation="MTMILSILSLFFIIRLLFLAVSIKHEKALIAKGAOKYGTNSTL LAATHLYLACFVWVWLSDTAFNGISLGTITVMASFVLSLITKOLGEIWTVKIY LPNHQINSWLEKFTFRHPNYFLNIPELIGIALLCQAWYVLLIGLPYLLVLEKRIQ EQQAWATLF"

misc_feature

49786..49809

/gene="NMA1102"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

complement(50214..50255)

repeat_unit

/note="Correia element; hmms hit to HMM Correia (114 - 156), score: 54.94"

repeat_unit

/label=Correia

50285..50344

/note="98% identity to consensus

TTCCGCCATCTCTAAGATTACAGCGATACACGGGTGATTTAAGGAATCCCCGAACC GTC"

repeat_unit

50345..50363
/note=">= 85% match to ATTCCNCNACNTTTCGTC"

/label=RS3x

repeat_unit

50364..50383
/note=">= 90% match to ATTCCNNNNNNNGGAAT"

/label=dRS3

repeat_unit

50384..50418

/note="100% identity to consensus

CTAGGACGCGAGGGTTAAGAAACCTACATCCGTC"

/label=RS38

repeat_unit

50419..50437
/note=">= 85% match to ATTCCNCNACNTTTCGTC"

/label=RS3x

repeat_unit

50438..50457

/note=">= 90% match to ATTCCNNNNNNNGGAAT"

/label=dRS3

repeat_unit

50458..50513

/note="100% identity to consensus

CTAGAACTCTCGGACTTTCAGATAATCTTTGAATATGCTGTTGTCTTAAGGTCCTAG"

/label=RS13

repeat_unit

50514..50533

/note=">= 90% match to ATTCCNNNNNNNGGAAT"

/label=dRS3

repeat_unit

50534..50603

/note="100% identity to consensus

GACGATTTCATAAGTTTCCGAAATTCACAACTAACCGAACCTGACAGTAACCGTAG CAACTGAACCGTC"

/label=RS16

repeat_unit

50604..50623

/note=">= 90% match to ATTCCNNNNNNNGGAAT"

/label=dRS3

repeat_unit

complement(50624..50679)

/note="82% identity to consensus

CCAGTCCGTTTCAGTTTCAGTCATTTCCGATAAATTCCTGTTGCTTTTCATTTCTAG"

/label=RS17

repeat_unit

50697..50850

/note="Correia element; hmms hit to HMM Correia (1 - 156), score: 258.97"

/label=Correia

50851..50886

/note="partial REP 3"

repeat_unit

complement(50887..51069)

/note="ATR repeat; hmms hit to HMM ATR (1 - 183), score: 308.46"

/label=ATR

51070..51096

/note="partial REP 3"

repeat_unit

51097..51115

/note=">= 85% match to ATTCCNCNACNTTTCGTC"

/label=RS3x

```

repeat_unit 51116..51135
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
repeat_unit 51136..51170
/notes="100% identity to consensus
CTAGGACGCGAGGTTAAGAAACCTACATCCCGTC"
/label=RS38
repeat_unit 51171..51190
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
repeat_unit 51191..51246
/notes="100% identity to consensus
CTAGAAATCTCGGACTTTCAGATAATCTTGAATATTCCTGTTCTTAAGGCTTAG"
/label=RS13
repeat_unit 51247..51266
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
repeat_unit 51267..51336
/notes="98% identity to consensus
GACGATTCATAGTTTCCCGAAATTCACATACCAACCAACCTGACAGTAACCGTAG
CACTGAACGGTC"
/label=RS16
repeat_unit 51337..51356
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
repeat_unit complement(51357..51412)
/notes="82% identity to consensus
CGAGTCGTCAGTTTCAGTCATTTCCGATAAATTCCTGCTTTTCATTTCTAG"
/label=RS17
repeat_unit 51430..51466
/notes="Correia element; hmms hit to HMM Correia (1 - 37),
score: 40.86"
/label=Correia
gene complement(51449..52036)
/genes="sodB"
CDS complement(51449..52036)
/genes="sodB"
/EC_number="1.15.1.1"
/notes="NMA1104, sodB, probable superoxide dismutase, len:
195 aa; highly similar to many e.g. SW:SODF_BORPE
(EMBL:X63055), sodF, Bordetella pertussis superoxide
dismutase [Fe] (EC 1.15.1.1) (192 aa), fasta scores: E():
0, 72.9% identity in 192 aa overlap. Contains Pfam match
to entry PF00081 sodF, Iron/manganese superoxide
dismutases (SODM) and PS00088 Manganese and iron
superoxide dismutases signature"
/transl_table=1
/product="putative superoxide dismutase"
/protein_id="CAB84366.1"
/db_xref="GI:7379798"
/translation="MEHKLQPLPYELDALSPHLSKETLEFFHYGKHQTYITNLNQIK
SFEFNLPLEETVKKSSGGVFNNAAQTNHTFYWLGFTEPKGQKPSGELAAAIADKAG
SPEKQEAFAACACTGCGWALVKTYPAGGLDLVSTNSAATPLTENTPLTCTDWE
HAYYIDYRNSRNYLKGFWEIWNDEAKRFAALS"
complement(51452..52033)
/genes="sodB"
/misc_feature /note="Pfam match to entry PF00081 sodF, Iron/manganese
superoxide dismutases (SODM), score 421.10, E-value
1e-122"
misc_feature complement(51545..51568)
/genes="sodB"
/notes="PS00088 Manganese and iron superoxide dismutases
signature"
RBS complement(52043..52047)
gene 52199..53605
/genes="dnaB"
CDS 52199..53605
/genes="dnaB"
/EC_number="3.6.1.-"
/notes="NMA1105, dnaB, probable replicative DNA helicase,
len: 468 aa; similar to many e.g. SW:DNAB_ECOLI

```

```

(EMBL:K01174), dnaB, Escherichia coli replicative DNA
helicase (EC 3.6.1.-) (471 aa), fasta scores: E(): 0,
54.2% identity in 459 aa overlap. Contains Pfam match to
entry PF00772 DnaB, DnaB-like helicase and PS00017
ATP/GTP-binding site motif A (P-loop). Contains probable
helix-turn-helix motif at aa 361-382 (Score 1036, +2.72
SD).
/codon_start=1
/transl_table=11
/product="putative replicative DNA helicase"
/protein_id="CAB84367.1"
/db_xref="GI:7379799"
/db_xref="SPTREMBL:Q9JUW8"
/translation="MNDYTAMPSEDEVGALSPLPHSMEAEQSVLGLMLLENPAWDRI
ADVSGEDFYRHEHRLIFRSIAKLINESRPADVTITVQEDLRNEELSAAGFEYLITL
AONTPSAANIRRYAEIVRSIMROLAEGVETARSAYNPQGRDAGQLDAEANKVFO
IAESTASKOGFLEMPDLLKEVVQIDMLYSRONPDVETGTCFIDLLKTSGLQPG
DLIIVAGRPSMGKTAFSINIAEHVAVEGKLPVAFVSEMGGAQLVMRLSGVGRDLOS
VLKTRLEDEHWRGLNEAVVKSADAPVYIDETPLGTALRLARRARLARFNNKLGIL
VIDYLQLMAGSGRSDNRSELSEISLAKALELQVPIIALSOLSPVESRTDKRPM
MSDLRESGAIEODADLIMFYRDEYYNQDSPMKGLAECLIIKHRNGSPVGKIFLTWTGQ
FTKFDNAAYIPEEAKIED"
52219..52228
/genes="dnaB"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
52331..53386
/genes="dnaB"
/notes="Pfam match to entry PF00772 DnaB, DnaB-like
helicase, score 551.30, E-value 6.5e-162"
52871..52894
/genes="dnaB"
/misc_feature /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
53732..53887
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 264.87"
/label=Correia
53914..54579
/genes="NMA1106"
53914..54579
/genes="NMA1106"
/notes="NMA1106, len: 221 aa; unknown, N-terminus is
similar to the N-termini of fimbrial proteins e.g.
TR:Q51528 (EMBL:L48934), fimT, Pseudomonas aeruginosa
fimbrial protein (169 aa), fasta scores: E(): 0.00046,
31.5% identity in 108 aa overlap. Contains possible
N-terminal signal sequence, similar to that present in
others from Neisseria meningitidis: NMA0264, NMA110,
NMA0726, NMA0424, which are all pilin-related"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1106"
/protein_id="CAB84368.1"
/db_xref="GI:7379800"
/db_xref="SPTREMBL:Q9JUW7"
/translation="MCTRKQQTTELLIVMVIAAVMAMIALPNMSQWTASRRIASH
AERIANLRFSGEAVRLNLPVICPVQVKDGTNNKCDGKGGKGLAFGDKNGNK
TVDGAADVFLRSVNLVNDTDDSRINFAFNHIAFGSSQPTADRVVTFENQGTGK
OHLTSKSSFFSDGYIQIIVLTARAVSDADKKFRSAVLINSSGRVEVCPNRDRVTC
QYK"
54609..55223
/genes="NMA1107"
54609..55223
/genes="NMA1107"
/notes="NMA1107, possible membrane protein, len: 204 aa;
unknown, N-terminus is similar to the N-terminus of
TR:Q57003 (EMBL:L36117), pilV, Pseudomonas aeruginosa gene
required for type 4 fimbrial biogenesis (185 aa), fasta
scores: E(): 0.00046, 25.1% identity in 183 aa overlap.
Contains hydrophobic, possible membrane-spanning region
near the N-terminus"
/codon_start=1
/transl_table=11

```



```
/label=Correia
61180..61959
/genes="NMA1114"
CDS
61180..61959
/genes="NMA1114"
/notes="NMA1114, len: 259 aa; similar to bacterial
hypothetical proteins e.g. SW:YAAA_ECOLI (EMBL:D10483),
yaaA, Escherichia coli hypothetical protein (258 aa),
fasta scores; E(): 0, 54.9% identity in 257 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1114"
/protein_id="CAB84376.1"
/db_xref="GI:7379809"
/db_xref="SPTREMBL:Q9JUV9"
/translation="MFFVLSPAKNLEKDPAPVSEFTOPDLAESDILMQOLRELAPO
QIAELMHVSDKIALNAQRNAENWTFEPENAKQAVFMENGDVYEGMDANTLNDIQ
YIGRVLLSLGLYGLRLDILQIOPYREMGTFANLRGNLYEFWGDIIINLNDLIA
QAGSNTLVNLAQSEYFKSVNTKKRLRITPIFKDEKNGKYKTIISFYAKRARGLMVRY
AAEHNTDPEMLKNFYEGYAFNDAASNESEWVFMRSQIK"
62042..62132
/notes="trna Ser anticodon TGA, Cove score 69.59"
complement(62201..62338)
/notes="REP 2; hmms hit to HMM REP 2 (1 - 128), score:
150.65"
trna
62345..62434
/notes="trna Ser anticodon CGA, Cove score 64.96"
62483..63698
/notes="IS1106A3"
repeat_unit
62483..62500
/notes="inverted repeat at end of IS1106A3"
/label=IR
complement(62630..63637)
/genes="NMA1115"
CDS
63994..64013
/notes="NMA1115, probable transposase for IS1106A3, len:
335 aa; similar to many e.g. TR:CAB44967 (EMBL:AJ242841),
Neisseria meningitidis transposase for insertion element
IS1106A3 (335 aa), fasta scores; E(): 0, 97.6% identity in
335 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transposase for IS1106A3"
/protein_id="CAB84377.1"
/db_xref="GI:7379809"
/db_xref="SPTREMBL:Q9JUV8"
/translation="MSTFFRQTAQMIAKHIDFPPLKLDQVIDWQPIEQYLNQRORT
YLDRHRGPAYPLLSMFKAVLLQWHLSLSPLEHSLITRIDNFLFCRFDLSIPDYS
TLCRYRNWLAODDTLSLELLKLNCLQTEKGLKTEKASAAVVDATITQTAGSKOROATE
VDEGOISGOTTPSKDSDARWIKKGLYKLYKQHTRTDAEGYIEKLYITPTNAHECT
HLSPLGLELPKGTIVYADKGYDAENRQHLERHQLQDGMKKACRNRPLESEVOTKRRN
A"
complement(63681..63698)
/notes="inverted repeat at end of IS1106A3"
/label=IR
63819..63973
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 258.02"
/label=Correia
complement(63974..63993)
repeat_unit
/notes="100% identity to consensus GAGCGGGACAGGGTTCCTGT"
63994..64013
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
/label=drs3
64014..64069
/notes="99% identity to consensus
CGAGTCGGTTCAGTTTCAGTTCATTCGATAAATTCCTGTGTTTTCATTCTAG"
/label=RS17
64070..64089
repeat_unit
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
```

```
/label=drs3
64090..64145
/notes="97% identity to consensus
GACGGCGGAGGGTTTGGTTTTTCGATAAATCTTTGAGGCATTGAAATTCAG"
/label=RS18
64146..64165
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
/label=drs3
64166..64235
/notes="100% identity to consensus
GAGATTTCATAAGTTTCCCGAAATTCACACATACCGAAACCTTCACAGTAACCGTAG
CAACTGAACCGTC"
/label=RS16
64236..64255
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
/label=drs3
64256..64311
/notes="88% identity to consensus
CTAGAATCTCGGAGCTTTTCAGATAATCTTTGAATATTGCTGTGTTCTTAAGGTCTAG"
/label=RS13
64312..64331
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
/label=drs3
64332..64437
/notes="94% identity to consensus
GACGAATCCATCCGACGGAAGCTGCACCACGTCATTCCTAGGAACCTACATCCCG
TCATTCCACACAGGACAGAAACCAAAATCAGAAACCTAAATCCCGTC"
/label=RS14
64437..64456
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
/label=drs3
complement(64457..64512)
/notes="94% identity to consensus
CGAGTCGGTTCAGTTTCAGTTCAGTTCGATAAATTCCTGTGTTTTCATTCTAG"
/label=RS17
64513..64532
/notes=">= 90% match to ATTTCCNNNNNNNGGGAAT"
/label=drs3
complement(64533..64558)
/notes="98% identity to consensus
AAGACCGTCGGGCATCTGCAGCGGTC"
64559..64707
/notes="Correia element; hmms hit to HMM Correia (1 -
151), score: 253.72"
/label=Correia
64749..64787
/notes="Stem loop containing DNA uptake sequences: ctaat
gcgctctgaa cgacgaatg ttcagacggc atttt"
64754..64763
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(64773..64782)
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(64799..67024)
/genes="icd"
complement(64799..67024)
/genes="icd"
/EC_number="1.1.1.42"
/notes="NMA1116, icd, probable isocitrate dehydrogenase,
len: 741 aa; similar to many e.g. SW:IDH_AZOV1
(EMBL:D73443), icd, Azotobacter vinelandii isocitrate
dehydrogenase (EC 1.1.1.42) (741 aa), fasta scores; E():
0, 65.6% identity in 735 aa overlap"
/codon_start=1
/transl_table=11
/product="putative isocitrate dehydrogenase"
/protein_id="CAB84378.1"
/db_xref="GI:7379810"
/db_xref="SPTREMBL:Q9JUV7"
/translation="MTQKSTIVYTHTDAPALATQSLPIVQAFARHADIDVKTADIS
LSGRILAAFPPEYLTEAORVPDALAEGLGVKQPDANVILKPNISASVPQLTAAIKELQ
SKGFAVPDYPADPQTDEKAVRERYDRIKGSAVNPVLREGNSDRRAPKAVKNFAKNP
```

```

HSMGAWTKDSKTHVATLQSGDFFHNEQSVTVDPDAISVSIIVTDTKQGNKKELRAPVAMK
AGEIIDATVMSKKALLAEFLAEQVQDAKAGVLFSLHMKATMMKVSDPIIFGHAVKAFK
APVEKEFGKGLAAGVNVNGFGLNLANLDKLDADTRAADVAAETAAVYAANPDLAMVD
SDKGIITNLHVPDSVIDVDSAMPATIRNSGRMWDKDKAQDTKAVIPDSSSYAGVQATID
FCRHGAFDPTTGCTVPGVGLMAQAABEYGSNNKTFFIEADGQVOVIDAAGKVLQMDH
VEAGDIWRMCQTADAPVKDQVQLAVNRTLSNTPAVFWDENRPHDKSLIAKVKAYILA
ELDTNGDIIRVLPAEEAKFSGLRKNEDTISVTGNVDRDYLTDLPFLELGTSAKM
LSIVPLNGMGFMFTGAGGAPKHVOQFLEENHLRDSLGEFLALAVSFEHLAOKTGN
AKAQVLNDTIDAATKLLNDKSPKAGELDNRGSFYLTLYWAQELAAQDKDAELK
AAPTPLAALTADAELKIVAELSAVQGAADIGGYAANPEKAQGMPSVTFNOALNA
L"
67057. .67254
/gene="NMA1117"
CDS 67057. .67254
/gene="NMA1117"
/notes="NMA1117, len: 65 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1117"
/protein_id="CAB84379.1"
/db_xref="GI:7379811"
/db_xref="SPTREMBL:O9JUV6"
/translation="MEHARGMCAAINAYFGGFVRALFDLGVFERGVKGRKWFSRV
WRSEVLWCMGGIFSFVKLLI"
complement(67295. .67356)
/notes="Correia element; hmms hit to HMM Correia (1 - 61),
score: 62.95"
/label="Correia"
67443. .67446
67457. .68572
/gene="lst"
67457. .68572
/gene="lst"
/EC_number="2.4.99.-"
/notes="NMA1118, lst,
CMP-N-acetylneuraminate-beta-galactosamide-
alpha-2,3-sialyltransferase, len: 371 aa; highly similar
to sialyltransferases from other strains e.g. TR:P72097
(EMBL:U60661), lst, Neisseria meningitidis strains 406Y,
NRCC and 4030 CMP-N-acetylneuraminate-beta-galactosamide-
alpha-2,3-sialyltransferase (EC 2.4.99.-) (371 aa), fasta
scores; E(): 0, 98.9% identity in 371 aa overlap"
/codon_start=1
/transl_table=11
/product="CMP-N-acetylneuraminate-beta-galactosamide-
alpha-2,3-sialyltransferase"
/protein_id="CAB84380.1"
/db_xref="GI:7379812"
/db_xref="SPTREMBL:O9JUV5"
/translation="MGLKKAFLVLCILVFCGIFVTFDRVNGERNVSLSKDKLFN
EEGEPVNLIFCYTILQMKVAERIMAGHPGERFYVILMSENRNEKYDYFNOIKDKAER
AYFFHLPYGLNKSFNFTPTMAELKVKSMLLPKVKRIYLASLEKVSAAFLSTYPDAAEI
KTFDDGTGNLIQSSVYLGDFFSVNGTKIKRNFARMIGDWSIAKTRNASDEHYTIFKGL
KNIMDGRKRMTVPLFDASELKAGDETGTGTVRILGSPDKEMKEISEKAANKFNQY
VAPHPQRTGLSGVTLSNIPYVIEDYILREIKNPNHRYEITFFFSGAALTMKDFPNV
HVTALPASLUPSDYWLKVPVYALFTQSGPIPLTFDDKD"
68690. .68722
/notes="Stem loop containing DNA uptake sequences: aaaa
gccgtctgaa gcc ttacagcgc attgt"
68695. .68704
68695. .68704
/notes="Core DNA uptake sequence: gccgtctgaa"
/label="DUS"
complement(68708. .68717)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label="DUS"
complement(68736. .69194)
/gene="NMA1119"
complement(68736. .69194)
/notes="NMA1119, probable C-type cytochrome, len: 152 aa;
similar to many e.g. SW:C554_RHOSH (EMBL:L36880), cycF,
Rhodobacter sphaeroides cytochrome C-554 precursor (153
aa), fasta scores; E(): 9.9e-06, 28.4% identity in 148 aa

```

```

overlap. Contains Pfam match to entry PF01322
Cytochrome_C_2, Cytochrome C', PS00190 Cytochrome c family
heme-binding site signature and an appropriately
positioned PS00013 Prokaryotic membrane lipoprotein lipid
attachment site. Contains a probable N-terminal signal
sequence"
/codon_start=1
/transl_table=11
/product="putative C-type cytochrome"
/protein_id="CAB84381.1"
/db_xref="GI:7379813"
/db_xref="SPTREMBL:O9JUV4"
/translation="MKTQISLAAAATLLLSACGGSGIPSQKGEISENRTAAFKSMX
PDFSRMGKMKVKEEYDVEKFKQAAAFAPESKKPFTLFESDPGNGRALPAVWSDGA
KFEAEKTKFAAAYEKLNAQAQTGKLDKKAAYGETGASCSDHSFRAPE"
complement(68742. .69107)
/gene="NMA1119"
/notes="Pfam match to entry PF01322 Cytochrome_C_2,
Cytochrome C', score 113.40, E-value 4.2e-30"
complement(68757. .68774)
/gene="NMA1119"
/notes="PS00190 Cytochrome c family heme-binding site
signature"
complement(69138. .69170)
/gene="NMA1119"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(69205. .69208)
69250. .69259
/notes="Core DNA uptake sequence: gccgtctgaa"
/label="DUS"
69303. .69306
69313. .69861
/gene="NMA1120"
/notes="Pfam match to entry PF00106 adh_short, short chain
dehydrogenase, score 191.70, E-value 1.2e-53"
69313. .70134
/gene="NMA1120"
69313. .70134
/notes="NMA1120"
/notes="NMA1120, probable oxidoreductase, len: 273 aa;
similar to many oxidoreductases e.g. TR:O54197
(EMBL:AJ000671), car, Streptomyces clavuligerus
clavulanate-9-aldehyde reductase (247 aa), fasta scores;
E(): 2.1e-23, 36.6% identity in 238 aa overlap and
SW:FABG_ECOLI (EMBL:M84991), fabG, Escherichia coli
3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)
(244 aa), fasta scores; E(): 1.1e-16, 32.8% identity in
201 aa overlap. Contains Pfam match to entry PF00106
adh_short, short chain dehydrogenase and PS00061
Short-chain dehydrogenases/reductases family signature"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAB84382.1"
/db_xref="GI:7379814"
/db_xref="SPTREMBL:O9JUV3"
/translation="MAVLITGASAGFGEAMCRAFGVGVIGAAARRADRLQALADEFL
GALFYPLEMDVSRRESVENALGIPDEFSDIDLNNAGLALGLDADKADFDEWETM
IQTNVLGLTLRTKLLPQWVERGGVYVNLGSIAGNAYAGSNVYGATKAFVROFSLN
LRAELADKNIRVTNIIEPLGCGNTEFSNVRFKGDDEKAAGYEGVEFTRPEDIAETALM
LYRRPAMNVNTIEIMPVQATFAGMKVIKAVPEVREDFEKQKSLFSRIRSWFK"
69712. .69798
/gene="NMA1120"
/notes="PS00061 Short-chain dehydrogenases/reductases
family signature"
70138. .70173
/notes="Stem loop containing DNA uptake sequences: ggaat
gccgtctgaa gacagt ttacagcgc atttt"
70143. .70152
/notes="Core DNA uptake sequence: gccgtctgaa"
/label="DUS"
complement(70159. .70168)

```

```

/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
gene      complement(70185. .70631)
          /gene="NMA1121"
CDS       complement(70185. .70631)
          /gene="NMA1121"
/note="NMA1121, possible acyl-CoA hydrolase, len: 148 aa;
shows weak similarity to part of SW:CTE2_RAT (EMBL:U49694)
Rattus norvegicus cytosolic acyl coenzyme a thioester
hydrolase (EC 3.1.2.2) (338 aa), fasta scores; E():
4.2e-07, 3l.0% identity in 113 aa overlap. Also similar to
hypothetical proteins e.g. SW:YC1A_HAEIN (EMBL:U32765),
Hi0827, Haemophilus influenzae hypothetical protein (154
aa), fasta scores; E(): 0, 62.1% identity in 132 aa
overlap. Similar to NMA1691, fasta scores; E(): 1.2e-08,
27.6% identity in 127 aa overlap. Contains pfam match to
entry PF01662 Acyl-CoA_hydro, Cytosolic long-chain
acyl-CoA thioester hydrolase"
/transl_table=1
/misc_feature complement(70185. .70631)
          /protein_id="CAB84383.1"
          /db_xref="GI:7379815"
          /db_xref="SPTREMBL:Q9JUV2"
          /translation="MQHEEGNRQPGQELLRLTVAMPRTDNPNDIFGWTMSQMDLG
GGILAAEIRGRIVTVAVQNMNFIRPVKGVNVVCYGHGCVRGVNTSLQLKVEVWVKTL
MNCVTEDRLVTEAVFTYVAIDAEGNRPRIKPGGNPILAGLLTP"
          /complement(70212. .70631)
          /gene="NMA1121"
/note="Pfam match to entry PF01662 Acyl-CoA_hydro,
Cytosolic long-chain acyl-CoA thioester hydrolase, score
160.80, E-value 2.3e-44"
RBS       complement(70638. .70641)
repeat_unit complement(70657. .70811)
          /note="Correia element; hnmfs hit to HMM Correia (1 -
156), score: 271.97"
/label=Correia
RBS       70903. .70906
gene      70913. .71845
          /gene="pip"
CDS       70913. .71845
          /gene="pip"
          /EC_number="3.4.11.5"
/note="NMA1122, pip, probable proline iminopeptidase, len:
310 aa; similar to many e.g. SW:PIP_NEIGO (EMBL:Z25461),
pip, Neisseria gonorrhoeae proline iminopeptidase (EC
3.4.11.5) (310 aa), fasta scores; E(): 0, 96.5% identity
in 310 aa overlap. Contains Pfam match to entry PF00561
abhydrolase, alpha/beta hydrolase fold"
/codon_start=1
/transl_table=1
/misc_feature complement(70185. .70631)
          /product="putative proline iminopeptidase"
          /protein_id="CAB84384.1"
          /db_xref="GI:7379816"
          /db_xref="SPTREMBL:Q9JUV1"
          /translation="MYEIKQPFHSGYLQVSEIHQIYWEESGNDPGVPVIFLHGPGGAG
ASPCRCGFENDPFRIVIIDRGCGSRPVACADNTTWDLVADIEKVRMLGIGKWL
VFGSGWSTLSLAYAQTHPERPVKGLVLRGFLCRPSETVNLNAGGVSRIYDEQWQX
VAPTAENRNQLIETAYHGLLFPHQDEEVLCSAAKAWDWSYLIRFEPPEVEDAYASL
AIARLENVFVNGLQGDRAILNNIGIKIHIPTIIVQGRYDLCTPMQSAWLSKAFP
EAELRVQAQGHRAFPDPLVDALVQAVEDILPHLL"
          /misc_feature complement(71084. .71839)
          /gene="pip"
/note="Pfam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 164.40, E-value 1.9e-45"
gene      complement(71902. .73029)
          /gene="NMA1123"
CDS       complement(71902. .73029)
          /gene="NMA1123"
/note="NMA1123, possible lipoprotein, len: 375 aa;
unknown, contains an appropriately positioned PS00013
Prokaryotic membrane lipoprotein lipid attachment site and
a possible N-terminal signal sequence"

```

```

/codon_start=1
/transl_table=1
/product="putative lipoprotein"
/protein_id="CAB84385.1"
/db_xref="GI:7379817"
/db_xref="SPTREMBL:Q9JUVO"
/translation="MTHIKPVIAALALIGLAACSGSKTEQPKLDYQSRSHRLIKLEVP
PDLNPPQGNLYRLPAGSAGVRAVDLEKRRTPAVQOPADAELVLSKVGRLERDGSOR
WLVVDGKSHAEIWPLLKAFWQENGFDLIKSEEPALGQMETWAENRAKIPQDSLRLFFD
TVGJGGIVYSTGERDKFIVRIEQGNVSDIFFAHKAMKEVYGGKDDTVMQPSPSDP
NLEAAFTREMOYLGVDGQAEASAKPPTLPAANEMARIEGKSLIVFGDYGRNWRRT
ALADRLGLTVGGQNTTERHAFLVQAGAPNESNAVTEQKPLGKRLRLLGKGAKPKPAEQE
LIVVAEPVANGSRIVLLNKDGSAYAGKADASALLKGLHSEL"
/misc_feature complement(71293. .73005)
          /gene="NMA1123"
          /note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
          /complement(73037. .73040)
          /complement(73101)
RBS       73069. .73101
stem_loop complement(73037. .73040)
          /note="Stem loop containing DNA uptake sequences: gttat
gccgtccgaa cgg ttcggacgc atggc"
          /complement(73107. .73982)
          /gene="dapa"
CDS       complement(73107. .73982)
          /gene="dapa"
          /EC_number="4.2.1.52"
          /note="NMA1124, dapa, probable dihydrodipicolinate
synthase, len: 291 aa; similar to many e.g. SW:DAPA_ECOLI
(EMBL:M12844), dapa, Escherichia coli dihydrodipicolinate
synthase (EC 4.2.1.52) (292 aa), fasta scores; E(): 0,
49.3% identity in 292 aa overlap. Contains pfam match to
entry PF00701 DHDPS, Dihydrodipicolinate synthetase
family, PS00665 Dihydrodipicolinate synthetase signature 1
and PS00666 Dihydrodipicolinate synthetase signature 2"
/codon_start=1
/transl_table=1
/product="putative dihydrodipicolinate synthase"
/protein_id="CAB84386.1"
/db_xref="GI:7379818"
/db_xref="SPTREMBL:Q9JU9"
/translation="MLQGSVALITPNNQDGSIIHQELRLDWHIENGTOGIVAVGT
TGESATLSVEEHTAVIEAVKVKAKRPVVIAGTGANNTVEALISQAQKAGADYTL
VVPYINPQSGMYRHFHKAEEAAIPMLYVPGRTVYSNNNETILRAELPIRVG
KEASNTGSLNIELINRAPEGVGDDHTALPFLMCGGHGIVTVAANAAPKLADMC
RAALQGDIALARELNDRLPIYDTMFCEPSPAAPKAWVSALGRCPEHVRLLPLVLPTEG
QAKYRAALKASGQL"
/misc_feature complement(73164. .73919)
          /gene="dapa"
          /note="Pfam match to entry PF00701 DHDPS,
Dihydrodipicolinate synthetase family, score 426.60,
E-value 2.3e-124"
          /misc_feature complement(74203. .74213)
          /note="Core DNA uptake sequence: gccgtctgaa"
          /label=DUS
          /complement(74280. .74283)
          /complement(74290. .75681)
          /gene="NMA1126"
          /complement(74290. .75681)
          /gene="NMA1126"
/note="NMA1126, probable transmembrane transport protein,
len: 463 aa; similar to many from both prokaryotes and
eukaryotes e.g. SW:UAPC_EMBNI (EMBL:X79796), uapc,
Emericella nidulans purine permease (580 aa), fasta
scores; E(): 4.4e-22, 30.7% identity in 499 aa overlap.
Also similar to hypothetical proteins e.g. SW:tfco_ECOLI
(EMBL:U28375), ygfO, Escherichia coli hypothetical protein
(485 aa), fasta scores; E(): 0, 69.4% identity in 438 aa
overlap. Contains Pfam match to entry PF00860
xan_ur_permease, Xanthine/uracil permeases family and
PS01116 Xanthine/uracil permeases family signature.
Contains hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=1

```

```

/product="putative transmembrane transport protein"
/protein_id="CAB84387.1"
/db_xref="GI:7379819"
/db_xref="SPTREMBL:Q9J025"
/translotion="MSGOLGKGADAPDLVVCLEDRPPFGNALLSAVTHLLAIFVPMILT
PALIVGALPELVEMTAYLYSYMNAVSGVTYQLQVNRFGPVGSGMLSIQSVNFSFVTV
MIALGAKKGGULFKDAMIISLLGVSVFGAFLVCFSAWLLPYLKKVITPIVSGVVMGL
IGLSLVHVGITDGGGFGKADGTGSMENGLASLLVLLVFNCKMKNPLLRSMVGA
VGLIAGIVALEFLGKVDFAQLONPLVTLVPVFKYGFADWHAFIVAGAIFLLSVFEA
VGLDTATAMYSODPIEGEYTKRLRGVLADGLSVIATAGLSPLPTFAQNGNCVIGM
TGVASRVGVKVIIVLLVLLGLFPVVGRAFTTIPSPVLGGAMVLMFGLIATAGVRLIYS
EAGRRRAVTAATSVGLGLGVAFEPVEFKMLPVLQFNSISAGGITAVALLMLVLPEDKT
EAAVFDTDHLEH"
/label=DUS
/gene="NMA1126"
/EC_number="74362. .75546"
/notes="Pfam match to entry PF00860 xan_ur_permease,
xanthine/uracil permeases family, score 436.60, E-value
2.1e-127"
/misc_feature
75385. .75447
/gene="NMA1126"
/notes="PS01116 Xanthine/uracil permeases family signature"
75691. .75733
/notes="Stem loop containing DNA uptake sequences: cgaat
gcgcgtcgaa acagaatccctgt ttcaagcggc attgt"
75696. .75705
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(75719. .75728)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(75743. .76528)
/gene="NMA1127"
/complement(75743. .76528)
/gene="NMA1127"
/notes="NMA1127, possible RNA methylase, len: 261 aa: shows
weak similarity to SW:TSNR_STRAZ (EMBL:X02392), tsnR,
Streptomyces azureus rRNA
(adenosine-2'-O-)-methyltransferase (EC 2.1.1.66) (269
aa), blastp scores; Expect = 4.0e-14. Similar to other
hypothetical RNA methylases e.g. SW:YJFH_ECOLI
(EMBL:U14003), Yjfh, Escherichia coli hypothetical
tRNA/tRNA methyltransferase (243 aa), fasta scores; E():
0, 52.3% identity in 241 aa overlap. Contains Pfam match
to entry PF00588 SpoU_methylase, SpoU rRNA Methylase
family"
/codon_start=1
/transl_table=11
/product="putative RNA methylase"
/protein_id="CAB84388.1"
/db_xref="GI:7379820"
/db_xref="SPTREMBL:Q9JU08"
/translotion="MPSEPFQSQT IMANORPIYGFHAVNARLWONPKSIVELYTOEGK
SDARTREVLKAAENIRVVFADADRLNAISKGARHOGVVGFIDASKNHVHLEDVLEN
LSEPLLILLIDGITDPHNLGACLRITADAMGVHAVIAPKDSAGLNATVSKVSGAET
VPIVTYNLARELURELKEYIGIWIIGDMGNADLIHCDLPDLSAAWVNGEGDMRRLLT
REHCDMLVSIPIMFCTGVESMVSVSAGMVLSETRQRRLKNEKA"
/complement(75785. .76210)
/gene="NMA1127"
/notes="Pfam match to entry PF00588 SpoU_methylase, SpoU
rRNA Methylase family, score 188.10, E-value 1.4e-52"
/complement(76517. .76526)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/label="NMA1127"
/gene="NMA1127"
/EC_number="76548. .76557"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(76562. .77050)
/gene="NMA1128"
/complement(76562. .77050)
/gene="NMA1128"
/notes="NMA1128, possible membrane protein, len: 162 aa;
similar to SW:YQCC_BAC50 (EMBL:D84432), YqgC, Bacillus

```

```

subtilis hypothetical protein (160 aa), fasta scores; E():
3.6e-10, 34.2% identity in 161 aa overlap. Contains
hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84389.1"
/db_xref="GI:7379821"
/db_xref="SPTREMBL:Q9JU07"
/translotion="MTALLVILALALTAAGTAGIVYPALPCLALMEFAGTWLLAYSGGV
QIYVAGVLTWVGLISLAGILADYVAGIWKYTGAGKLAVRGALAGSLIGFISLPGL
ILGFFGAAAGELIERNNLQAGKAGLGTLLGLLVYFAFKIGCAVSILFILLVXIAY
LF"
/complement(77055. .77501)
/gene="NMA1129"
/complement(77055. .77501)
/gene="NMA1129"
/EC_number="3.5.4.1"
/notes="NMA1129, possible cytosine deaminase, len: 239 aa;
shows weak similarity to many e.g. SW:FCAL_CANAL
(EMBL:U55194), fcal, Candida albicans cytosine deaminase
(EC 3.5.4.1) (150 aa), fasta scores; E(): 5.5e-08, 31.4%
identity in 140 aa overlap. Contains Pfam match to entry
PF00383 dCMP_cyt_deam, Cytidine and deoxycytidylate
deaminase zinc-binding region and PS00903 Cytidine and
deoxycytidylate deaminases zinc-binding region signature"
/codon_start=1
/transl_table=11
/product="putative cytosine deaminase"
/protein_id="CAB84390.1"
/db_xref="GI:7379822"
/db_xref="SPTREMBL:Q9JU06"
/translotion="MEHFMCALRQAEQSAADGEIPVGAVIVSDGKIITASAHNTCIAD
CNVSRHAEINALAQAGREMONYRLDGDYITLPPCMACASALIOARVKRYVYGAGEP
KTGAAGSIVNLFADKRLNTHTAIRGGILQEBCEAVLSRFFQNKRRG"
/complement(77235. .77354)
/gene="NMA1129"
/notes="Pfam match to entry PF00383 dCMP_cyt_deam, Cytidine
and deoxycytidylate deaminase zinc-binding region, score
67.20, E-value 3.9e-17"
77409. .77418
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(77504. .77509)
/complement(77544. .77762)
/gene="NMA1129a"
/complement(77544. .77762)
/gene="NMA1129a"
/notes="NMA1129a, possible pseudogene, len: 219 bp; shows
weak similarity to part of SW:TFOX_HAEIN (EMBL:U13205),
tfox, Haemophilus influenzae DNA transformation protein
(217 aa), fasta scores; E(): 0.0005, 38.4% identity in 73
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1229a (pseudogene)"
/complement(77778. .78719)
/gene="miaA"
/complement(77778. .78719)
/gene="miaA"
/EC_number="2.5.1.8"
/notes="NMA1130, miaA, probable tRNA
delta(2)-isopentenylpyrophosphate transferase, len: 313
aa; similar to many e.g. SW:MTAA_ECOLI (EMBL:M63655),
miaA, Escherichia coli tRNA
delta(2)-isopentenylpyrophosphate transferase (EC 2.5.1.8)
(316 aa), fasta scores; E(): 0, 52.1% identity in 280 aa
overlap. Contains PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=11
/product="putative tRNA delta(2)-isopentenylpyrophosphate

```

```

transferase"
/protein_id="CAB84392.1"
/db_xref="GI:7379823"
/db_xref="SPTREMBL:Q9JUJ5"
/translation="MPTPKAFALLGPTAGGKTALALKIETLPEVILSLSALVYRDM
DIGTAPSAERAFVPHLLIDIITPVQTSAAKFVEDCTRLIGETIARGKPLIVGGT
MMYFRALTOGLNDLPADACLRDLDEQKOMYGLDFLYRTLQOVDPETACRLKPNDSQ
RIGRALEVYLTGKPMSEHLGRQSPHTLPFDLHTAALIPENRAHLHENIALRFLHMLE
QGIQGVENLRRYPSSLTADSPAIRCVCVYQAWWEYLDGATDRQTFTEKGIATRQLAK
RLTLRLKTPPLDCVADPFSDGTSCTRLIEAKREFGV"
7829..77838
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(78256..78265)
/genes="miaA"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(78666..78689)
/genes="miaA"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(78831..78985)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 273.06"
/label=Correia
complement(79137..79146)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
79191..79499
/genes="NMA1131"
79191..79499
/genes="NMA1131"
/notes="NMA1131, possible membrane protein, len: 102 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84393.1"
/db_xref="GI:7379824"
/db_xref="SPTREMBL:Q9JUJ4"
/transl_table=11
/product="MAFCFFESLANSRLVRLVCLIPLATLAVFAANPPEDKPKQH
LINGIILACATFLFKFVLFDTIKHLLKQEFDLKQKTMLLFIPIILLIVLYFHYFGAF
"
79518..80666
/genes="NMA1132"
79518..80666
/genes="NMA1132"
/notes="NMA1132, len: 382 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1132"
/protein_id="CAB84394.1"
/db_xref="GI:7379825"
/db_xref="SPTREMBL:Q9JUJ3"
/translation="MNTPPFVCVIFCKVIDNFGDIGVSWRLARVLRHRELGMQVHLWTD
DVSAIRALCPDLDPVCPVHQDIIHVRTWHSDAADIDTAPVDPVVIETFACDLPENLHI
IRHKPLVLNWEYLSAESNERLHLMPSQESQVKYFWFMFGSEKSGLLIRERDYCEA
VRDGSALRLRLPEKNAPEWLLFLGVSQWAKLWEMWQAGSPLTLLLAGAILDS
LKQNGVTPQALQNDGDVQFQTSVRLVKIIFVQQDFDKLLHLADCAVIRGEDSFVRA
QLAGKPFWHIYPQDENVHLDKLHAFWDKAHGFTTPTASAHRLRLSDDLNGGEALSAT
QRLECWQILQHQHNGWQGAEDWSRYLFGQPSASEKLAAFVSKHQKTR"
complement(80227..80236)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
80709..81266
/genes="efp"
/notes="Pfam match to entry PF01132 EFP, Elongation factor
P (EF-P), score 192.00, E-value 9.4e-54"
80709..81269
/genes="efp"
80709..81269
/genes="efp"
/notes="NMA1133, efp, probable elongation factor P, len:

```

```

186 aa; similar to many e.g. SW:EFP_ECOLI (EMBL:X61676),
efp, Escherichia coli elongation factor P (187 aa), fasta
scores: E(): 1e-18, 35.5% identity in 183 aa overlap.
Contains Pfam match to entry PF01132 EFP, Elongation
factor P (EF-P)"
/codon_start=1
/transl_table=11
/product="putative elongation factor P"
/protein_id="CAB84395.1"
/db_xref="GI:7379826"
/db_xref="SPTREMBL:Q9JUJ2"
/translation="MKTAQELRAGNVFMVGNPVMVQKTEYIKGGRSAKVSMLKML
LTGAASETIYKADDFKDFVILSRKNCTYSYFADPMYFMDEEFNQYIEADNIGDALUK
FIVDMEDQCEVTFEYGNPISVELPTIIVREVETEPVAVGDTSGKVMKTKARLVGGTPE
IQVMSYIENGDKIEIDTRTGEFRKA"
81283..81317
/notes="Stem loop containing DNA uptake sequences: aaaat
gccgtctgaa aacgt ttcacagcgc atttt"
81288..81297
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(81303..81312)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(81319..82164)
/genes="NMA1134"
complement(81319..82164)
/genes="NMA1134"
/notes="NMA1134, possible lipoprotein, len: 281 aa;
unknown, contains appropriately positioned PS00013
Prokaryotic membrane lipoprotein lipid attachment site and
probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAB84396.1"
/db_xref="GI:7379827"
/db_xref="SPTREMBL:Q9JUJ1"
/translation="MKNKTSSLMLLAAMMLTACSPSKEDTKENGASAASTASAAS
SSAPQTDLPAAASAPDNVKNQAESVPPSNCTDLHPATGIDDLMOQIAEHIDSDCLFALS
HMELETRFGLPGGQDNIQRLLFPDIPEDPDYHQKIIALEDYRKRTISRQAQDA
LMEOERLRERATLLLIQGSQETQGGEPKRTYFEVSATPAYSRRNNNGGNNFYII
QQLPGYLIKHGEMLENGSLFSLNRERNPKPFLDIHFDENGKITRIVVYEKNIFYNP
NLGRR"
complement(82105..82137)
/genes="NMA1134"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
82189..82198
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(82216..82225)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(82281..82862)
/genes="NMA1135"
complement(82281..82862)
/genes="NMA1135"
/notes="NMA1135, len: 193 aa; unknown, shows weak
similarity to parts of hypothetical proteins e.g.
TR:032813 (EMBL:U80599) Lactococcus lactis hypothetical
protein (207 aa), fasta scores: E(): 0.001, 25.6% identity
in 121 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1135"
/protein_id="CAB84397.1"
/db_xref="GI:7379828"
/db_xref="SPTREMBL:Q9JUJ0"
/translation="MNLRDDLLIYDRIPEGSRVLDLGGCGDELLAASVEHKCSGYG
TEIDTDSVIAMSRGVNIQADLEEGLTAFNDQSFVIVLSQTIQAQNTKEKILRCLM
RVAKQALVSFNFYGVYNIQALGGHMPYSERPYHWYDTPNIHWCTKDFDLCLAK
NNIRVLRAVMTGNRQVKHFPNLLGSLAFYRG"

```



```

/note="NMA1141, possible redoxin, len: 245 aa; similar to
many from both bacteria and eukaryotes e.g. TR:AAD28242
(EMBL:AF121355) Arabidopsis thaliana peroxiredoxin (162
aa), fasta scores: E(): 2e-17, 39.9% identity in 158 aa
overlap. Contains Pfam match to entry PF00578 AhpC-TSA,
AhpC/TSA family"
/codon_start=1
/transl_table=11
/product="putative redoxin"
/protein_id="CAB84403.1"
/db_xref="GI:7379834"
/translation="NALQDRTGQKVPVFRVRVGTWKDVSTDLFGKKVYVFSLP
GAFTCSSSHLPYRNELFGAFKENGDAIYGVNDTFVNMWAAEESDNIYIPD
NGEFTEGMLVGKEDLFGKRWYSMLVNDGVKMFIEPFGPGDPFKVSDATM
LOFVAPDKAQSVAIFTKPGCOFCAKAKQALQDKLSVEEIVLGDATVTSVRAITG
KWTAPQVFIGKYIGSDELEAYLAKN"
88809..89276
/genre="NMA1141"
/note="Pfam match to entry PF00578 AhpC-TSA, AhpC/TSA
family, score 157.90, E-value 1.7e-43"
89713..89722
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
89765..89768
RBS
89777..91180
gene
89777..91180
CDS
89777..91180
/genre="lpdA2"
/EC_number="1.8.1.4"
/note="NMA1142, lpdA2, probable dihydrolipoamide
dehydrogenase, len: 467 aa; similar to many e.g.
SW:DLD1_BACST (EMBL:X53560), pdhD, Bacillus
stearothermophilus dihydrolipoamide dehydrogenase (EC
1.8.1.4) (470 aa), fasta scores: E(): 0, 30.0% identity in
454 aa overlap. Similar to NMA1151, fasta scores: E():
3.5e-28, 28.6% identity in 482 aa overlap and NMA1556,
fasta scores: E(): 1.9e-24, 26.0% identity in 469 aa
overlap. Contains Pfam match to entry PF00070 pyr_redox,
Pyridine nucleotide-disulphide oxidoreductase class-I."
/codon_start=1
/transl_table=11
/product="putative dihydrolipoamide dehydrogenase"
/protein_id="CAB84404.1"
/db_xref="GI:7379835"
/db_xref="SPTREMBL:Q9JUT5"
/translation="MKKIQADVVIIGGTAGMCAFRNARLHSDNYLIENWVGITCA
RVGCMPSKLLIAAEARHHLHTDFGVHLDKDSIVVNGEVMQRKSDRDRFVGVV
ADGEMPADKRIMGSAKFIDEHTVQIDEHTQITAKSFVIATSRPVIIPQWOSLGNRL
IINDVFSWDTLPKRVAVFGPGVIGLELQALRLGVKVEIFGLGIGIIGSDPVVSD
EAKAVFGEELKLHLDAKTEAKLDAGNVHWEQDGEKGVFAEYMLAAVGRPNVD
IGLENIITEKDARGVPVADPLTMQTSIPHIFTAGDASNLPLLHEAADQGTAGDNAG
RYPNITGGILRRSTIGVFTSPQIGFVGLKYAQVQADFEVIGVSEFKNQGRSRVM
LVNKGHMLIYAEKATGRFICGAEIVGPAAEHLALHLAWAHQMKWTVPQMLDMPFIHPVI
EEGLRTAURDADAKLKA"
89798..91177
/genre="lpdA2"
/note="Pfam match to entry PF00070 pyr_redox, Pyridine
nucleotide-disulphide oxidoreductase class-I, score
111.00, E-value 2.3e-29"
91194..91228
stem_loop
/note="Stem loop containing DNA uptake sequences: acaat
gccgtctgaa atttt ttacagacgc atttt"
91199..91208
misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(91214..91223)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
91459..91836
gene
91459..91836
CDS
91459..91836
/genre="sdhC"

```

```

/note="NMA1143, sdhC, probable succinate dehydrogenase
cytochrome B subunit, len: 125 aa; similar to many e.g.
SW:DHSC_ECOLI (EMBL:J01619), sdhC, Escherichia coli
succinate dehydrogenase cytochrome B-556 subunit (129 aa),
fasta scores: E(): 1.4e-14, 39.8% identity in 123 aa
overlap. Contains Pfam match to entry PF01127 Sdh_cyt,
Succinate dehydrogenase cytochrome b subunit and PS01000
Succinate dehydrogenase cytochrome b subunit signature 1.
Contains hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative succinate dehydrogenase cytochrome B
subunit"
/protein_id="CAB84405.1"
/db_xref="GI:7379836"
/db_xref="SPTREMBL:Q9JRJO"
/translation="MSAKPRPVYLDLPNRLPIPIGVSTILHRISGVGLFIMLPFLLYF
LSGTLQSQSAFETYRAIVSHPLVLDLIGVLWAYLHLSLAGIRFLFLDAHKGLELNTA
RNTAKAVFASALVLTIVLGALLW"
91465..91815
/genre="sdhC"
/note="Pfam match to entry PF01127 Sdh_cyt, Succinate
dehydrogenase cytochrome b subunit, score 79.90, E-value
5.3e-20"
91474..91548
misc_feature
91474..91548
/genre="sdhC"
/note="PS01000 Succinate dehydrogenase cytochrome b
subunit signature 1"
91830..92171
gene
91830..92171
CDS
91830..92171
/genre="sdhD"
/genre="sdhD"
/note="NMA1144, sdhD, probable succinate dehydrogenase
hydrophobic membrane anchor protein, len: 113 aa; similar
to SW:DHSD_ECOLI (EMBL:J01619), sdhD, Escherichia coli
succinate dehydrogenase hydrophobic membrane anchor
protein (115 aa), fasta scores: E(): 5e-13, 36.5% identity
in 104 aa overlap. Also similar to TR:O83000
(EMBL:AB015757) Rhodofex fermentans fumarate reductase
cytochrome B small subunit (121 aa), fasta scores: E():
1.8e-21, 48.6% identity in 111 aa overlap. Contains
hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative succinate dehydrogenase hydrophobic
membrane anchor protein"
/protein_id="CAB84406.1"
/db_xref="GI:7379837"
/db_xref="SPTREMBL:Q9JUT4"
/translation="MYERKLTGAHYGLRDWAMQRATAVIMLIYTVALLVYVLPALPREY
SAWQAFFSQTVKVTQVSFIADVFLHAWYGIKPGVRLFLQVATIVMLVG
CLVYSIKVTWG"
92174..93937
gene
92174..93937
CDS
92174..93937
/genre="sdhA"
/genre="sdhA"
/EC_number="1.3.99.1"
/note="NMA1145, sdhA, probable succinate dehydrogenase
flavoprotein subunit, len: 587 aa; similar to many e.g.
SW:DHSA_ECOLI (EMBL:J01619), sdhA, Escherichia coli
succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)
(588 aa), fasta scores: E(): 0, 56.6% identity in 590 aa
overlap. Also similar to TR:O83001 (EMBL:AB015757)
Rhodofex fermentans fumarate reductase flavoprotein
subunit (601 aa), fasta scores: E(): 0, 64.2% identity in
592 aa overlap. Similar to NMA2092, fasta scores: E():
8.6e-31, 33.9% identity in 546 aa overlap. Contains Pfam
match to entry PF00890 FAD_binding_2, FAD binding domain
and PS00504 Fumarate reductase / succinate dehydrogenase
FAD-binding site"
/codon_start=1
/transl_table=11
/product="putative succinate dehydrogenase flavoprotein

```



```

QAYRDALDKGEHVEQTTLNSFORTQIDWSKYQGKDWREHIETGLPAAIERLTERKFTA
VPEGFALPTAKRVITARKMAGSKQAIIDWGMATFLAYASLVTKGHGVRSIGESGGRG
TFSHRAHVLDHOKREKWDGTVVPLRHMGEGEFLVIDSLNEFAVNAFEGFACSA
PDKLTIRWQAGSDFEANGAQTIDQFLSSGETKWGLCGLTTLPHCYDQGGPSSAR
VERWLOLCSGNMVMPSSEASQMFHLQRQVLGSRYPPLVFMPSKRLRLRPFKGMSP
ENFTGSGTFRPVIGDTABRASNDSKRVVLVCAGQVYYDLEAGRAERKLEDDVAIVRVE
QLYPPYDEVKAEALAKYFNAKSVVMAQEPKNQGAQFYQIRHIEDVISEEQKLSYAGR
PSSASPAYGYSKSHIAQLKQLVEDALAL"
/misc_feature
97234. .98214
/genes="sucA"
/EC_number="2.3.1.61"
/EC_number="2.3.1.61"
/note="NMA1150, sucB, probable dihydrolipoamide
succinyltransferase E2 component, len: 403 aa; similar to
many E2 components of 2-oxoglutarate dehydrogenase complex
e.g. SW.ODO2_ECOLI (EMBL:J01619), sucB, Escherichia coli
dihydrolipoamide succinyltransferase component of
2-oxoglutarate dehydrogenase complex (EC 2.3.1.61) (404
aa), fasta scores; E(): 0, 57.1% identity in 408 aa
overlap. Similar to NMA1555, fasta scores; E(): 8.9e-24,
35.8% identity in 433 aa overlap. Contains Pfam matches to
TAAACAPAAA repeats at aa 79-113. Contains Pfam matches to
entry PF00198 2-oxoacid_dh, 2-oxo acid dehydrogenases
acyltransferase (catalytic domain) and to entry PF00364
biotin_lipoyl, Biotin-requiring enzymes and PS00189 2-oxo
acid dehydrogenases acyltransferase component lipoyl
binding site"
/codon_start=1
/transl_table=11
/product="putative dihydrolipoamide succinyltransferase E2
component"
/protein_id="CAB84412.1"
/db_xref="GI:7379843"
/db_xref="SPTREMBL:Q9JUT2"
/translation="MIDVKVPMLSSEVSGTLLKWKVKVGAVARDEILIDIEDTKV
VLEVPSPQAGVLVEIQAQGETVVAQDQVLARIDTAATAAAEAPAAATAAAEAPAA
EAPAAAPAAQNAAMPAAAKLAAESGVDVNALQSGRGRVLKEDVDONAAKPAAP
AAPAVLPAGARPEERVPMSRLRARVAERLLASQOENALTTFNEVNMKPIMDLRKY
KEKFEKHGVKLGFMSFFVKAARALKKYPVNVNASVDGKDVIYHGYFDIGIAGSPRG
LVVPILRADQMSIADIEQAIVDYAKKADGKIDLELTGFTGFTSYNGFTGSGMSTP
LIINPQSAILGMHATKERAUVENGQVVRPMMYLAISDYDHRDIIDGREAVLTILVAIKDA
LEDPARLLIDL"
/misc_feature
99550. .99771
/genes="sucB"
/note="Pfam match to entry PF00364 biotin_lipoyl,
Biotin-requiring enzymes, score 102.80, E-value 6.6e-27"
99622. .99711
/genes="sucB"
/note="PS00189 2-oxo acid dehydrogenases acyltransferase
component lipoyl binding site"
/misc_feature
100090. .100749
/genes="sucB"
/note="Pfam match to entry PF00198 2-oxoacid_dh, 2-oxo
acid dehydrogenases acyltransferase (catalytic domain),
score 420.10, E-value 2e-122"
/misc_feature
complement(100760. .100769)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
100791. .100800
/note="Core DNA uptake sequence: gccgtctgaa"

```

```

/label=DUS
complement(100805. .100847)
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 50.38"
/label=Correia
complement(100849. .100911)
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 78.71"
/label=Correia
complement(101011. .101020)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
101109. .101112
/genes="lpdA3"
101120. .102553
/genes="lpdA3"
101120. .102553
/genes="lpdA3"
/EC_number="1.8.1.4"
/note="NMA1151, lpdA3, probable dihydrolipoamide
dehydrogenase E3 component, len: 477 aa; similar to many
E3 components of 2-oxoglutarate dehydrogenase complex e.g.
SW:DLDH_ALCEU (EMBL:X91877), odhL, Alcaligenes eutrophus
dihydrolipoamide dehydrogenase (EC 1.8.1.4) (474 aa),
fasta scores; E(): 0, 62.3% identity in 472 aa overlap.
Similar to NMA1556, fasta scores; E(): 0, 38.2% identity
in 476 aa overlap and NMA1142, fasta scores; E(): 5.5e-26,
28.9% identity in 481 aa overlap. Contains Pfam match to
entry PF00070 pyr_redox, Pyridine nucleotide-disulphide
oxidoreductase class-I and PS00076 Pyridine
nucleotide-disulphide oxidoreductases class-I active site"
/codon_start=1
/transl_table=11
/product="putative dihydrolipoamide dehydrogenase E3
component"
/protein_id="CAB84413.1"
/db_xref="GI:7379844"
/db_xref="SPTREMBL:Q9JUT1"
/translation="MSQYDVVVGAGPGGYVAAIRAAQGLFKTACVDAGVKNAGNAPA
LGCTCLNVGCIPSKALQSSEHFHAAQHFHITGVGVDFVDVAKMIERKDAIVTKL
TGGVFLQKQKVKVTSIFCTPASFAGKNGDAYOIEVONKGEKTVLEAKHVTAVTGSVPRP
LPQVAIDNVLDNEGALNLTEVPKALGVIGSGVIGSEMGSVNVRVGAETILEARPT
FLAADDQQIAKEAFYFTKEQGLSTELGVKIDIKSEMGSVVAVETAAGEARKEVFD
KLIVAGIRIPNTKGLNAAEAVGLEKDERGFKVDCERTNLPNVMAIGDVVRGPMIAHG
ASDEGVAAERIAQGRHIDFNNVPFIYTDPEIAWVGTEQLKAEQVEYKKGTSGF
GANGRALAMGAKGTVKVLADAKTDRLILGVHMGIPVWSELLTEGVTALEFFASSEDIA
RIIHAHPTLSEVYHEAALAADKRALHG"
101135. .102550
/genes="lpdA3"
/note="Pfam match to entry PF00070 pyr_redox, Pyridine
nucleotide-disulphide oxidoreductase class-I, score
680.90, E-value 6.3e-201"
101255. .101287
/genes="lpdA3"
/note="PS00076 Pyridine nucleotide-disulphide
oxidoreductases class-I active site"
102648. .102935
/genes="NMA1152"
102648. .102935
/genes="NMA1152"
/note="NMA1152, possible membrane protein, len: 95 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions and a possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84414.1"
/db_xref="GI:7379845"
/db_xref="SPTREMBL:Q9JR20"
/translation="MNKEIVGIFFPAGLIISMCAALWQYVMYETVTLNRFKDEL
VWRVALLFISFLAVYLLICPNRKKGVIFVILGGGAAYLLARMLWLPFSK"
103023. .103027
103034. .104200
/genes="sucC"

```

RBS
gene

CDS

```

103034. .104200
/ gene="sucC"
/ EC-number="6.2.1.5"
/ note="NMA1153, sucC, probable succinyl-CoA synthetase
beta subunit, len: 388 aa; similar to many e.g.
SW:SUCC_ECOLI (EMBL:J01619), sucC, Escherichia coli
succinyl-CoA synthetase beta chain (EC 6.2.1.5) (388 aa),
fasta scores: E(): 0, 60.5% identity in 387 aa overlap.
Contains pfam match to entry PF00549 ligase-CoA,
CoA-ligases and PS01217 ATP-citrate lyase / succinyl-CoA
ligases family signature 3"
/ codon_start=1
/ transl_table=11
/ product="putative succinyl-CoA synthetase beta subunit"
/ protein_id="CAB84415.1"
/ db_xref="GI:7379846"
/ db_xref="SPTREMBL:Q9JUT0"
/ translation="MNLREYQAKELLASYGLPVQGGTILAHNGEAAAYDKLGKFAV
VKAQVHAGGRGKAGGVKVKVRSREAEKVAESLGTNLVTVYTANGOPVNSVLCEDEM
YPOTELYLGAVVDRSPRTVFNASTEGGVEIEKVAETPEKIFKVTVDPLVGLQPCQ
AREVAFQLGKDKQINEFYKLMTGAYKAFVENDFALFEVNPNAVRENGALACVDGKIG
IDSNALRLPKIAELRDKSQENERELEKASEFDLNYVALEGNIGCMVNGAGLAMATMDI
IKLGGQANFLVDGGGATKDRVVEAFKLLEDKSVKGLVINIFGGIVRCDMAEATV
AAVKEINVDVPVVVRLGEGNNAELGAKILNESGLKLTADGLNDAAEKIVAANA"
103772. .104191
/ gene="sucC"
/ note="Pfam match to entry PF00549 ligase-CoA,
CoA-ligases, score 259.70, E-value 3.8e-74"
103805. .103879
/ gene="sucC"
/ note="PS01217 ATP-citrate lyase / succinyl-CoA ligases
family signature 3"
104200. .104204
/ gene="sucC"
104211. .105101
/ gene="sucD"
104211. .105101
/ gene="sucD"
/ EC-number="6.2.1.5"
/ note="NMA1154, sucD, probable probable succinyl-CoA
synthetase alpha subunit, len: 296 aa; highly similar to
many e.g. SW:SUCD_ECOLI (EMBL:J01619), sucD, Escherichia
coli succinyl-CoA synthetase alpha chain (EC 6.2.1.5) (288
aa), fasta scores: E(): 0, 72.1% identity in 290 aa
overlap. Contains pfam match to entry PF00549 ligase-CoA,
CoA-ligases, PS01216 ATP-citrate lyase / succinyl-CoA
ligases family signature 1 and PS00399 ATP-citrate lyase /
succinyl-CoA ligases family active site"
/ codon_start=1
/ transl_table=11
/ product="putative succinyl-CoA synthetase alpha subunit"
/ protein_id="CAB84416.1"
/ db_xref="GI:7379847"
/ db_xref="SPTREMBL:Q9JUS9"
/ translation="MSVLINKDTKVLVQGTGKNGTFHSEQALAYCTKVVGWTPGKG
GQTHLDLPVFTMKKAVKETGADASVIYVPAPEVLDSIVEAVDSGVLVVITEGVTPT
LDMKAKRYLETNGNCTRLVGNPCVITPGECKIGIMPGLHITPGRGIISRSGLTL
YEAVAQTKLGIQSTCGIGSGDPIPGMNIQIDALKALFQEDPOTDALIMGEIGGTAEE
EAERYIQSNVSPVVGVIAGVTPAPKGRMGHAGAIISGKGTAEEKFAFEKAGIAYT
RSPAELGTMLLEVLTKGLA"
104613. .105056
/ gene="sucD"
/ note="Pfam match to entry PF00549 ligase-CoA,
CoA-ligases, score 210.80, E-value 2.1e-59"
104670. .104759
/ gene="sucD"
/ note="PS01216 ATP-citrate lyase / succinyl-CoA ligases
family signature 1"
104922. .104963
/ gene="sucD"
/ note="PS00399 ATP-citrate lyase / succinyl-CoA ligases
family active site"
105127. .105170
/ note="Stem loop containing DNA uptake sequences: caaat

```

misc_feature

misc_feature

gene

CDS

```

gccgtctgaa accggaatcggtggt ttcagacggc atttt"
105132. .105141
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(105156. .105165)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(105336. .105698)
/ gene="NMA1155"
/ gene="NMA1155"
/ note="NMA1155, len: 120 aa; unknown, shows similarity to
parts of hypothetical proteins e.g. SW:YAFM_HAEIN
(EMBL:U032706), HI0217, Haemophilus influenzae hypothetical
protein (176 aa), fasta scores: E(): 2.6e-15, 42.5%
identity in 106 aa overlap. Similar to parts of NMA1675,
fasta scores: E(): 8.4e-07, 42.7% identity in 103 aa
overlap and NMA0532, fasta scores: E(): 8.4e-07, 42.7%
identity in 103 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein NMA1155"
/ protein_id="CAB84417.1"
/ db_xref="GI:7379848"
/ db_xref="SPTREMBL:Q9JUS8"
/ translation="MSNISASRAAYMDVQKQYPTETVAVCVLPNHIIHAIWTLPPDDAD
YSLRLRIKTKFSAYSPTTKNLGAKQQRHGIWQRFRFGHTVRDETDLQRCADYIH
SNPIKQMSRAKTESCKK"
105944. .106174
/ gene="NMA1156"
/ gene="NMA1156"
105944. .106174
/ gene="NMA1156"
/ note="NMA1156, len: 76 aa; unknown, lysine-rich"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein NMA1156"
/ protein_id="CAB84418.1"
/ db_xref="GI:7379849"
/ db_xref="SPTREMBL:Q9JUS7"
/ translation="MWRSHARGVGYAGYGLLLQKLLKKPKPTPLKRLKKKQKTKRKM
KLLKTKLKKLNNSDSFTEKRIPIGNCNPFVL"
complement(106211. .106220)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
106231. .106240
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(106340. .107137)
/ gene="NMA1157"
complement(106340. .107137)
/ note="NMA1157, len: 265 aa; unknown, lies within a region
of unusually low GC content"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein NMA1157"
/ protein_id="CAB84419.1"
/ db_xref="GI:7379850"
/ db_xref="SPTREMBL:Q9JUS6"
/ translation="MQAFOISPAVATRDPAFYCSFSKNLRDFTKSKFYLTSTEDFEPQ
FPHSFENAIKILNGSLIFDISQHQPLLPFDLPVLPINPDVLIKSKALSLLKKPLK
IYHSITSSGLSEKVIPLAVLNDGFRWHLRFDKTKRNFADFVLRITIOANILLAEIN
EENETINNDIDWNRIVKTELVPNGHDEAKKVIEMDFMGKGLKLEALRASNIQVYLQ
QWRVDCSSDHHIDDPAFRLWLLDLNLVLYGINNAHLAPAFDKKRKP"
complement(107861. .108049)
/ gene="NMA1158"
complement(107861. .108049)
/ gene="NMA1158"
/ note="NMA1158, len: 62 aa; unknown, lies within a region
of unusually low GC content"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein NMA1158"

```



```

/misc_feature
len: 213 bp, similar to parts of transposases e.g.
TR:Q48208 (EMBL:X58176) Haemophilus influenzae insertion
sequence IS1016(V-4) ORF2 (164 aa), fasta scores; E():
1.1e-09, 61.2% identity in 49 aa overlap. Similar to many
others from Neisseria meningitidis e.g. NMA1543, fasta
scores; E(): 1.8e-07, 62.9% identity in 35 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative transposase (pseudogene)"
118523..119557
/gene="NMA1167"
118523..119557
/gene="NMA1167"
/notes="NMA1167, possible phage protein, len: 344 aa; shows
weak similarity to e.g. TR:O34630 (EMBL:U83796), rsta2,
Vibrio cholerae bacteriophage CTXphi protein required for
replication and integration (359 aa), fasta scores; E():
1.5e-09, 27.8% identity in 223 aa overlap. Also similar to
TR:O82959 (EMBL:AB015669) ORF C7 from Burkholderia
solanacearum plasmid pUTPS1 (444 aa), fasta scores; E():
1.4e-16, 33.9% identity in 239 aa overlap. Highly similar
to others from Neisseria meningitidis, NMA1999, E(): 0,
100.0% identity in 344 aa overlap, NMA0782, fasta scores;
E(): 0, 42.7% identity in 337 aa overlap, and NMA1792,
fasta scores; E(): 0, 43.3% identity in 298 aa overlap.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/transl_table=11
/product="putative phage protein"
/protein_id="CAB84429.1"
/db_xref="GI:7379859"
/db_xref="SPTREMBL:Q9JR20"
/translation="MEDEIFRAASMLAEVFGFGIYKESKSGGRFYERCLWLGSD
VLGRVHFQGNNTILFELTGCVGAKGWSRLFAFLNALPKRITRVDIARDFFN
GYSPNQAEEDRNKMTFCHHVKPGECIDGSDWEEDDEAKMTKTYGICRSSESKY
RVEYKQKQDKTSTWTRFEIEFKADIVIPFVLQNPGEYFGGAYICRFYQKATP
IHAVKEDVTSADRYLWVKKQFGRAANGKLFPPDLDAKLFELIPSHHKLPKSLA
PEAYDCAFLKAQAIHEQPAFKPKYKDPYMYEYENLEKLEQEQKHVNNEESYNNFYD
KFARLPSIWA"
118523..121495
/notes="region containing CDS similar to NMA0776-NMA0782,
NMA1792-NMA1797 and NMA1999-NMA2005"
118943..118966
/gene="NMA1167"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
119548..119892
/gene="NMA1168"
119548..119892
/gene="NMA1168"
/notes="NMA1168, len: 114 aa; unknown, similar to others
from Neisseria meningitidis, NMA2000, fasta scores; E():
0, 100.0% identity in 114 aa overlap, NMA1793, fasta
scores; E(): 1.6e-13, 37.8% identity in 98 aa overlap and
NMA0781, fasta scores; E(): 2.3e-10, 34.9% identity in 83
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1168"
/protein_id="CAB84430.1"
/db_xref="GI:7379860"
/db_xref="SPTREMBL:Q9JS46"
/translation="MGLKCLPERRLLITQGNKNKNIQLQGHIVGVKKFNGOIEGRSPDY
CLRIVATPLDSSQGNALGSSTTEYDFGGSANFQFRNAQFPIEANLVEIVTTGKTQK
LKVIGFQPVKKG"
119897..120094
/gene="NMA1169"
119897..120094
/gene="NMA1169"
/notes="NMA1169, len: 65 aa; unknown, similar to others
from Neisseria meningitidis, NMA2001, fasta scores; E():
4.2e-27, 100.0% identity in 65 aa overlap, NMA1794, fasta
scores; E(): 4.6e-07, 38.3% identity in 60 aa overlap and

```

```

/misc_feature
len: 213 bp, similar to parts of transposases e.g.
TR:Q48208 (EMBL:X58176) Haemophilus influenzae insertion
sequence IS1016(V-4) ORF2 (164 aa), fasta scores; E():
1.1e-09, 61.2% identity in 49 aa overlap. Similar to many
others from Neisseria meningitidis e.g. NMA1543, fasta
scores; E(): 1.8e-07, 62.9% identity in 35 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative transposase (pseudogene)"
118523..119557
/gene="NMA1167"
118523..119557
/gene="NMA1167"
/notes="NMA1167, possible phage protein, len: 344 aa; shows
weak similarity to e.g. TR:O34630 (EMBL:U83796), rsta2,
Vibrio cholerae bacteriophage CTXphi protein required for
replication and integration (359 aa), fasta scores; E():
1.5e-09, 27.8% identity in 223 aa overlap. Also similar to
TR:O82959 (EMBL:AB015669) ORF C7 from Burkholderia
solanacearum plasmid pUTPS1 (444 aa), fasta scores; E():
1.4e-16, 33.9% identity in 239 aa overlap. Highly similar
to others from Neisseria meningitidis, NMA1999, E(): 0,
100.0% identity in 344 aa overlap, NMA0782, fasta scores;
E(): 0, 42.7% identity in 337 aa overlap, and NMA1792,
fasta scores; E(): 0, 43.3% identity in 298 aa overlap.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/transl_table=11
/product="putative phage protein"
/protein_id="CAB84429.1"
/db_xref="GI:7379859"
/db_xref="SPTREMBL:Q9JR20"
/translation="MEDEIFRAASMLAEVFGFGIYKESKSGGRFYERCLWLGSD
VLGRVHFQGNNTILFELTGCVGAKGWSRLFAFLNALPKRITRVDIARDFFN
GYSPNQAEEDRNKMTFCHHVKPGECIDGSDWEEDDEAKMTKTYGICRSSESKY
RVEYKQKQDKTSTWTRFEIEFKADIVIPFVLQNPGEYFGGAYICRFYQKATP
IHAVKEDVTSADRYLWVKKQFGRAANGKLFPPDLDAKLFELIPSHHKLPKSLA
PEAYDCAFLKAQAIHEQPAFKPKYKDPYMYEYENLEKLEQEQKHVNNEESYNNFYD
KFARLPSIWA"
118523..121495
/notes="region containing CDS similar to NMA0776-NMA0782,
NMA1792-NMA1797 and NMA1999-NMA2005"
118943..118966
/gene="NMA1167"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
119548..119892
/gene="NMA1168"
119548..119892
/gene="NMA1168"
/notes="NMA1168, len: 114 aa; unknown, similar to others
from Neisseria meningitidis, NMA2000, fasta scores; E():
0, 100.0% identity in 114 aa overlap, NMA1793, fasta
scores; E(): 1.6e-13, 37.8% identity in 98 aa overlap and
NMA0781, fasta scores; E(): 2.3e-10, 34.9% identity in 83
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1168"
/protein_id="CAB84430.1"
/db_xref="GI:7379860"
/db_xref="SPTREMBL:Q9JS46"
/translation="MGLKCLPERRLLITQGNKNKNIQLQGHIVGVKKFNGOIEGRSPDY
CLRIVATPLDSSQGNALGSSTTEYDFGGSANFQFRNAQFPIEANLVEIVTTGKTQK
LKVIGFQPVKKG"
119897..120094
/gene="NMA1169"
119897..120094
/gene="NMA1169"
/notes="NMA1169, len: 65 aa; unknown, similar to others
from Neisseria meningitidis, NMA2001, fasta scores; E():
4.2e-27, 100.0% identity in 65 aa overlap, NMA1794, fasta
scores; E(): 4.6e-07, 38.3% identity in 60 aa overlap and

```

```

NMA0780, fasta scores; E(): 1.3e-06, 36.1% identity in 61
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1169"
/protein_id="CAB84431.1"
/db_xref="GI:7379861"
/db_xref="SPTREMBL:Q9JQ6"
/translacion="MQKYVYVSVSTGDFLYLSPETGDIHTKLTNADYFYDFEAI
120102..120386
/gene="NMA1170"
120102..120386
/gene="NMA1170"
/notes="NMA1170, possible integral membrane protein, len:
94 aa; unknown, similar to others from Neisseria
meningitidis, NMA2002, fasta scores; E(): 0, 100.0%
identity in 94 aa overlap. Contains hydrophobic, possible
membrane-spanning regions, NMA1170, fasta scores; E(): 0,
100.0% identity in 94 aa overlap and NMA0779, fasta
scores; E(): 1.5e-18, 72.2% identity in 72 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84432.1"
/db_xref="GI:7379862"
/db_xref="SPTREMBL:Q9JRX7"
/translacion="MFGRLKNAPSITAKHFLKENIMKFINCKRYGAKLAVVTAAPL
ALAAHANATLPDTAKNALEAAKADGMEAGVIVGVFAALFVFSIVKRVK"
120393..120671
/gene="NMA1171"
120393..120671
/gene="NMA1171"
/notes="NMA1171, len: 92 aa; unknown, similar to others
from Neisseria meningitidis, NMA2003, fasta scores; E():
0, 100.0% identity in 92 aa overlap, NMA0778, fasta
scores; E(): 1.5e-25, 58.7% identity in 92 aa overlap and
NMA1796, fasta scores; E(): 8.4e-09, 32.3% identity in 93
aa overlap. C-terminus is hydrophobic"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1171"
/protein_id="CAB84433.1"
/db_xref="GI:7379863"
/db_xref="SPTREMBL:Q9JRF3"
/translacion="MYQVGNKCLERHQENLYFSLVVPRIKENGQIVRPEYNSLWK
MSDQPLRLILAECSPKDNLQSLGTGWIIVFGLASVYFVSLKVLK"
120799..121116
/gene="NMA1172"
120799..121116
/gene="NMA1172"
/notes="NMA1172, len: 105 aa; similar to others from
Neisseria meningitidis e.g. TR:087782 (EMBL:AJ010115)
Neisseria meningitidis strain B:15:pl.16 hypothetical
protein (fragment) (148 aa), fasta scores; E(): 0, 83.8%
identity in 105 aa overlap. Also similar to NMA2004, fasta
scores; E(): 4e-25 54.8% identity in 104 aa overlap and
NMA0777, fasta scores; E(): 6.7e-16, 57.5% identity in 73
aa overlap. Lies within a region of unusually low GC
content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1172"
/protein_id="CAB84434.1"
/db_xref="GI:7379864"
/db_xref="SPTREMBL:Q9JUR8"
/translacion="MNKPFITQALYKYQSPSKYFGOSMALIAQKEFEFFVNNVKE
YDILESFTLKNRAHNLIWYFSDSEVIFIRKSENGKTVEHFVYQYTDSSDFNS
ME"
121058..121498
/gene="NMA1173"
121058..121498
/gene="NMA1173"

```

```

/notes="NMA1173, probable cell-surface protein, len: 146
aa; highly similar to the N-terminus of TR:087783
(EMBL:AJ010115), tsbp, Neisseria meningitidis strain
B:15:pl.16 T-cell and B-cell antigen (partial CDS) (504
aa), fasta scores; E(): 5.9e-29, 68.2% identity in 129 aa
overlap. Similar to others from Neisseria meningitidis,
e.g. NMA0776, fasta scores; E(): 1.8e-23, 51.0% identity
in 145 aa overlap, and NMA1797, fasta scores; E():
1.8e-23, 51.0% identity in 145 aa overlap (both tsbp
genes) and NMA2005 (another putative cell-surface
protein), fasta scores; E(): 1.2e-30, 66.2% identity in
133 aa overlap. N-terminal half has low GC content.
Contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative cell-surface protein"
/protein_id="CAB84435.1"
/db_xref="GI:7379865"
/db_xref="SPTREMBL:Q9JUR7"
/translacion="MNSFIKSTPIVLILICSSNSYSEPVRLERSQIKFQSSNNLKSS
GFKLDSSKSFAPKFIKFAANFEHIPTGAKARINAKITASVSRAVLGVGKLARLGAKF
STRAVPYVGTALLAHDVYETFKEDIQARGYQYDPTDKFAKVS"
121676..121694
/notes=">= 85% match to ATTCCCNACNNTTTCGTC"
/label=RS3x
121695..121714
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drS3
121715..121770
/notes="100% identity to consensus
CTAGAATCTCGACATTCAGATAATCTTGAATATTCCTGTGTTCTTAAGTCTTAG"
121771..121790
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drS3
complement(121791..121841)
/notes="98% identity to consensus
GACGGCGTAAGAGGTTCTGTGTTTTTAAGTTTAGTAACTAATGAGTCGTC"
/label=RS62
121842..121861
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drS3
complement(121862..121917)
/notes="98% identity to consensus
CCAATTCGTTCGGTTTCAGCCGTTCCGATAAATCACTGTACGTTAAGTTCCCGG"
/label=RS41
121918..121937
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drS3
121938..121996
/notes="99% identity to consensus
GACGAATCCATCCGTACGGAACCTGCACCACGTCATCCACGAACCTACATCCCG
TC"
/label=RS14a
121997..122016
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drS3
122017..122077
/notes="95% identity to consensus
CTAGGACGTAAATCTAAGAAACCGTTTACCCGATAAGTTTCCGACCGACGACAGAC
CTAG"
/label=RS15
122078..122097
/notes=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drS3
122102..122569
/gene="NMA1174"
122102..122569
/gene="NMA1174"
/notes="NMA1174, len: 155 aa; unknown, similar, except at
the N-terminus, to SW:MDAB_ECOLI (EMBL:U18656), mdab,
Escherichia coli modulator of drug activity B (putative
modulator of topoisomerase IV activity imparting DMP 840,

```

adriamycin and etoposide resistance on overexpression)
(193 aa), fasta scores; E(): 0, 63.7% identity in 146 aa
overlap. Also similar to hypothetical proteins from
bacteria and Schizosaccharomyces pombe e.g. SW:YA05_SCHPO
(EMBL:Z49811), SPAC5H10.05C, Schizosaccharomyces pombe
hypothetical protein (196 aa), fasta scores; E(): 0, 55.1%
identity in 147 aa overlap. Identical, except at the
N-terminus, to NMA0600, fasta scores; E(): 0, 99.3%
identity in 147 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1174"
/protein_id="CAB84436.1"
/db_xref="GI:7379866"
/db_xref="SPTREMBL:O9JUR6"
/translation="MOVFLTIPGYDEAEIEKFEVMDAVIWMQPMHEPWTKKYI
DGVLTAGHKLYQSDGRHSNPTGEGTGGLLOGKHKMLSLTNWAPIEAETREGDFE
GKGVDLYIMHFHRADEFGLMTRLPTEFLCNDVVKNPQVEKYLADYQAHLEKVF"
122657..122666
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
122685..122731
complement(122766..122775)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(122777..124162)
/gene="pntB"
complement(122777..124162)
/gene="pntB"
complement(122777..124162)
/EC_number="1.6.1.1"
/note="NMA1175, pntB, probable NAD(P) transhydrogenase
beta subunit, len: 461 aa; similar to many e.g.
SW:PNTB_ECOLI (EMBL:X04195), pntB, Escherichia coli NAD(P)
transhydrogenase beta subunit (EC 1.6.1.1) (462 aa), fasta
scores; E(): 0, 77.3% identity in 463 aa overlap. Contains
hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative NAD(P) transhydrogenase beta subunit"
/protein_id="CAB84437.1"
/db_xref="GI:7379867"
/db_xref="SPTREMBL:O9JUR5"
/translation="MSSGLTAAIYIVAAIILFISLAGLSKQETAKGCGSYGIAGMAVA
LFYVFSNDTHGLGWIIIAMLGAAIYKAKKRVEMTEPILALHSHFVGLAAVLVG
FNSYIESGNVSHMDTHLVEVVLGTFIGAVTFGSLVAFGKNGKITSPPLOLPKAKH
KNALAAVSFVLVLFVFGINGSFTILLIMTLIALAFGHLVASIGGADMPVVSMLN
SYSGWAAARGFMLSNDLLIVTALVSGSGAILSYIMCKAMNRSFVSIVAGFGSDSS
AASFAGEVGEYREVRAADVAEMLKGANSVLIITPGYGMVAQAQYPPVAETELLKNG
TEYRFGIHPVAGRLPGHMNVLLAEKVPYDIVLEMDIINDDFETDVLVIGANDTVN
PAAQTPNPISPIAGMPVLEVMKAEVVFVKRSMNTGYAGVQNPLFFNENSVCMCFGDAKK
TVDDIILSELKK"
complement(124170..124173)
124198..124207
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(124228..124611)
/gene="NMA1176"
complement(124228..124611)
/gene="NMA1176"
/note="NMA1176, possible membrane protein, len: 127 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84438.1"
/db_xref="GI:7379868"
/db_xref="SPTREMBL:O9JUR4"
/translation="MTFAVWCILIAVLLPLFCAYAKKAGGFRKONHNPRDFLARTQ
GTAARAHAAQQNGFEAFAPFAAAVLTAHATGNAGQATVNTLAGLFIILFRLAIFWCYIA
DKAALRLSMWVGGFVCTVGLFVVA"
complement(124615..126156)
/gene="pntA"

CDS
complement(124615..126156)
/gene="pntA"
/EC_number="1.6.1.1"
/note="NMA1177, pntA, probable NAD(P) transhydrogenase
alpha subunit, len: 513 aa; similar to many e.g.
SW:PNTA_ECOLI (EMBL:X04195), pntA, Escherichia coli NAD(P)
transhydrogenase alpha subunit (EC 1.6.1.1) (510 aa),
fasta scores; E(): 0, 69.8% identity in 513 aa overlap.
Contains pfam match to entry PF01262 AlaDh_PNT, Alanine
dehydrogenase/pyridine nucleotide transhydrogenase and
PS00837 Alanine dehydrogenase & pyridine nucleotide
transhydrogenase signature 2. Contains hydrophobic,
probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative NAD(P) transhydrogenase alpha subunit"
/protein_id="CAB84439.1"
/db_xref="GI:7379869"
/db_xref="SPTREMBL:O9JUR3"
/translation="MKIGIPRESISGSTRVACTPATVALLKGLGFFETVSGAGLAAS
LDDAAVQAAGATVADKAAVWAYPLIYKVNAPSEDELPLLEGOTIVSFLEWPRONEALV
EALRAKKNALAMDMPRISSAQALDALSSMANTISGYRAVIEAANAGREFTGQITAA
GKVPFAQVLVIGAGVAGLAAGTANSIGAVVRAFDRLEVAEQIESMGGFELKLDFFQ
ESGSGDGYAKVMSDEFIAAEKMLFAEQAKEVDIIITTAIPGKPAKPLKITKEMVSM
KPGSVIVDLAATGCGELTKQGELEFVTGTGVKIIIGYTMANRLAGSSQLYATNLVN
LTKLLSPNKDGEITLDPEDVIIRNMTVRDGEITFPDPPIQVSAQPOQTPSEKAAPAA
KPKPKVPLWKKLAPVIAAVLVWGVAPAAFLNHFIVFLVACVIGYVYVWNVSHS
LHTPLMSVTNAISGIIIVGALLQIGQGNGFVSLLSFVAILLASINIFGGEFVTRMLN
MFRKG"
complement(124623..124626)
/gene="pntA"
complement(124990..126156)
/gene="pntA"
/note="Pfam match to entry PF01262 AlaDh_PNT, Alanine
dehydrogenase/pyridine nucleotide transhydrogenase, score
697.80, E-value 5.3e-206"
complement(124999..125008)
/gene="pntA"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(125572..125649)
/gene="pntA"
/note="PS00837 Alanine dehydrogenase & pyridine nucleotide
transhydrogenase signature 2"
complement(126164..126167)
complement(126284..126400)
/gene="NMA1178"
complement(126284..126400)
/gene="NMA1178"
/note="NMA1178, len: 38 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1178"
/protein_id="CAB84440.1"
/db_xref="GI:7379870"
/db_xref="SPTREMBL:O9JUR2"
/translation="MLRRAGLQSEPDFKRRQYVSDNTGKCRNYCQPDENSL"
126446..127279
/gene="serB"
126446..127279
/gene="serB"
/EC_number="3.1.3.3"
/note="NMA1179, serB, probable phosphoserine phosphatase,
len: 277 aa; similar, except at N-terminus, to many e.g.
SW:SERB_ECOLI (EMBL:X03046), serB, Escherichia coli
phosphoserine phosphatase (EC 3.1.3.3) (322 aa), fasta
scores; E(): 8.4e-30, 45.7% identity in 234 aa overlap"
/codon_start=1
/transl_table=11
/product="putative phosphoserine phosphatase"
/protein_id="CAB84441.1"
/db_xref="GI:7379871"
/db_xref="SPTREMBL:O9JUR1"

```

/translacion="MPQALVLPSPSAEALPSPDFSRPLPEPCADKRMRFIVEGFSL
SGADAALLGSCQIDHAVLPMDFGLGLIVSDMSDLTITIECVDEIAEGVLKNAVE
ITERSMRGELDEFOSLSRVALLAGLDEKILADYENVLKLSGCAFELDECKRHVK
FLVSGGFTEFTELQORLGFYFQHANILEIENGRLTGRLKRIIDAQAKADLLREYR
SLUGLGHQVHLAMGDGANDIPMLKEAGIGVAYRAKPKARAADACINFGGLERVGLF
G"
misc_feature
127038..127047
/genes="serB"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/label=DUS
127292..127328
/note="Stem loop containing DNA uptake sequences: cggat
gccgtctgaa aaacgtg ttcacagcgc atttg"
misc_feature
127297..127306
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
misc_feature
complement(127314..127323)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
127335..127338
/genes="NMA1180"
127345..127890
/genes="NMA1180"
127345..127890
/genes="NMA1180"
/note="NMA1180, probable integral membrane protein, len:
181 aa; shows weak similarity to the N-terminal half of
Escherichia coli hypothetical proteins e.g. TR:P77195
(EMBL:D90802) Escherichia coli hypothetical protein (430
aa), fasta scores; E(): 3.2e-05, 26.8% identity in 183 aa
overlap. Also shows weak similarity to the N-terminal half
of NMA0434, fasta scores; E(): 0.018, 27.7% identity in
119 aa overlap. The following CDS, NMA1181, is similar to
the C-termini of these proteins and is separated by a
[gcc]4 repeat. Contains possible N-terminal signal
sequence and hydrophobic, probable membrane-spanning
regions"
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84442.1"
/db_xref="GI:7379872"
/db_xref="SPTREMBL:Q9JU09"
/translacion="MGRKRLVLAALAAAGVIGLVGLVITLTMHFVQHTAYGYCAGCVY
TSFREGVQAQSMRRVAVLTLCGATAGGCGWLLKRGKQPIEKALKALOPLAGLPLT
TVFVLLQITLVGLSGPVGREVAQPREMTAAFAFAGGRIGLDEDEMRLLIACASGAGL
AAVYNVPLASTLFIILEAMLVG"
127915..128568
/genes="NMA1181"
127915..128568
/genes="NMA1181"
/note="NMA1181, probable integral membrane protein, len:
217 aa; shows weak similarity to the C-terminal half of
bacterial hypothetical proteins e.g. SW:YADQ.ECOLI
(EMBL:D26562), yadQ, Escherichia coli hypothetical protein
(436 aa), fasta scores; E(): 2.2e-05, 25.0% identity in
212 aa overlap. Also shows weak similarity to the
C-terminal half of NMA0434, fasta scores; E(): 7.2e-06,
25.6% identity in 211 aa overlap. The previous CDS,
NMA1180, is similar to the N-termini of these proteins and
is separated by a [gcc]4 repeat. Contains hydrophobic,
possible membrane-spanning regions"
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84443.1"
/db_xref="GI:7379873"
/db_xref="SPTREMBL:Q9JU09"
/translacion="MLTSVIATAVARIGLDVQOYHPANLAVNTSLMFSVAIGPLG
ATAVFRSAEKFFPLKRDNPRIIPALATFALICAIYVWFPEILGNKAGNQLTFGG
LTDWHSLELTAVKWLVLVLMALAAAGYGLITPSPMGLSTISFAAAANRWYFPEMP
ESAAVYGAVFGLVGSNLNPLTAAVFVLELTYPATALLPLCTGMAGAVVWTARKMGCK"
128388..128397
/genes="NMA1181"
misc_feature

```

```

/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
128577..128613
/note="Stem loop containing DNA uptake sequences: aaat
gccgtctgaa accggtt ttcacagcgc attcc"
128582..128591
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(128599..128608)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
128721..128724
128732..130324
/genes="purH"
128732..130324
/genes="purH"
/EC_number="2.1.1.2.3"
/EC_number="3.5.4.10"
/note="NMA1182, purH, probable bifunctional purine
biosynthesis protein, len: 530 aa; similar to many e.g.
SW:PUR9.ECOLI (EMBL:J05126), purH, Escherichia coli
bifunctional purine biosynthesis protein (EC 2.1.2.3) (EC
3.5.4.10) (529 aa), fasta scores; E(): 0, 59.5% identity
in 528 aa overlap. Probably encodes
phosphoribosylaminoimidazolecarboxamide formyltransferase
and IMP cyclohydrolase"
/transl_table=11
/product="putative bifunctional purine biosynthesis
protein"
/protein_id="CAB84444.1"
/db_xref="GI:7379874"
/db_xref="SPTREMBL:Q9JU08"
/translacion="MPSIKRALISLSDKTGAVEFAQTLHLGLVEILSTGGTAKLLADA
GVPIEVADVTGPEMLDGVKTLHPKIHGGILGRDLPEHVAKNKHWALVTDPAFPAIAEME
NLYPFAATLAKPNCLETEDATENTDIGPMTVRSAKNKHWALVTDPAFPAIAEME
ANNALSDDKTRENLSKAFSHTAQYDGMISNYLSISDDVLSGTPEIGEPSPQFNQSW
IKVDMMRYGENPHQRAAFYNDVTPAAGLSAYKOLGKELSYNNIADADAWEAKVSF
DAPACVIVKHANPCGVAAADTLTAYLATDTTSAFGGIIAFNRVDGDETQKQITD
NOFMEVLMADPKFTAALAAKKNRVRLQISLTTPLEAGANFELKRVGGGLVQTP
DIYLRNADLKVVSKRLTQEWNDLMFVWNAKYKSNVAFVGKGTQYIGAGOMS
RVDSTRIARKAODANLDLNGACAASDAFFPFRDGVDIVIAEQGIIKALIHPSGMRDQE
VDAADDEGHVAVYTVGRHFRH"
130339..130373
/note="Stem loop containing DNA uptake sequences: gtaat
gccgtctgaa gggct ttcacagcgc atttt"
130344..130353
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(130359..130368)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
130452..130456
130468..130692
/genes="NMA1183"
130468..130692
/genes="NMA1183"
/note="NMA1183, len: 74 aa; unknown"
/transl_table=11
/product="hypothetical protein NMA1183"
/protein_id="CAB84445.1"
/db_xref="GI:7379875"
/db_xref="SPTREMBL:Q9JU07"
/translacion="MGCRCVCFSPHQKFLFNRESAVLVSAGTVMGNIDGIDSGGL
KPTLVVNHCIITNRPAAATRAHPRTTVA"
complement(130798..130952)
/note="Correia element; hmms hits to HMM Correia (1
136), score: 240.56"
/label=Correia
130937..131890
/genes="NMA1185"
/note="possible IS element"

```


gene

CDS

130937. .131893
/gene="NMA1185"
130937. .131893
/gene="NMA1185"
/note="NMA1185, probable transposase, len: 318 aa; similar to many e.g. TR:051474 (EMBL:M2471) Alteromonas atlantica CDS within insertion element IS492 which is involved in reversible inactivation of extracellular polysaccharide production (318 aa), fasta scores; E(): 1.3e-28, 33.9% identity in 301 aa overlap. Also shows weak similarity to pilin-inverting genes e.g. SW:PIV_MORLA (EMBL:N34367), piv, Moraxella lacunata pilin gene inverting protein (322 aa), fasta scores; E(): 8.5e-14, 30.4% identity in 313 aa overlap. Highly similar to others from Neisseria meningitidis e.g. NMA0733, fasta scores; E(): 0, 94.3% identity in 279 aa overlap, also NMA2085. Contains Pfam match to entry PF01548 Transposase_9, Transposase"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="CAB84446.1"
/db_xref="GI:7379876"
/db_xref="SPTREMBL:O9JUQ6"
/translation="MLIHYIDIARKNFVIAVSSLSKTKTETNPNKGIATHTVEYLKKHK VALVVTSTGLEIPAAKARIAGVLIANPQTHQFAQSPLTKTDADKAKPAFF AQKAQEDQWTPYHPPTAEAEVLEALVRNRPADMTAEKNRLQVHETQVESYK QLTAFRLIDELDKQIDDDHTHFDGQAQVAEQIKGISTTTATLMLPELGRLSH KRATSLVGIAPHPRESGETKFKSRCFGRSVARKALYMATVAATREPELRIDFQRLP SEGKPYKAVYACMRKLLTISNARMRDYFAENDTAENGRTA"
131126. .131431
/gene="NMA1185"
/note="Pfam match to entry PF01548 Transposase_9, Transposase, score 34.10, E-value 3.3e-06"
131989. .132483
/gene="NMA1186"
131989. .136706
/note="possible partial prophage element pnm3"
/label=pnm3
131989. .132483
/gene="NMA1186"
/note="NMA1186, len: 164 aa; unknown, similar to part of TR:Q38494 (EMBL:M64097) bacteriophage Mu protein E16 (195 aa), fasta scores; E(): 2.7e-05, 46.7% identity in 60 aa overlap. Identical to NMA1301 and similar to NMA1867, fasta scores; E(): 1.5e-25, 55.5% identity in 137 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1186"
/protein_id="CAB84447.1"
/db_xref="GI:7379877"
/db_xref="SPTREMBL:O9JS31"
/translation="MYFIGICVNSIYCYVEPEMRRLAIKIAKQKELGLDGTVR AVLERVTDKSCADMVSELESVADMRSHGFKPKAGNPGHPHLRRTSSAAMLDKV EALLTVGGKHWNVAHAMARRMFGKDKVEYLDLDTQLHLKVAALQIAENRKTEKAGGDDG VRKS"

gene

CDS

132464. .132892
/gene="NMA1187"
132464. .132892
/gene="NMA1187"
/note="NMA1187, len: 142 aa; unknown, identical to NMA1302 and similar to NMA1866, fasta scores; E(): 2.4e-10, 29.7% identity in 145 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1187"
/protein_id="CAB84448.1"
/db_xref="GI:7379878"
/db_xref="SPTREMBL:O9JRW0"
/translation="MCFEKVHELLPDVLDIVDVLIGLAATEQLVKAIGCARFEGKCK VDERLAILVEATGEVKTHELLQVGGEEELYVPCGKALQLRNRHRYQEFVKLRDID KKSGLMAWTKLCPKYGISSTGYTIINEMSRPAQAALF"
complement(133017. .133171)

misc_feature

gene

repeat_unit

gene

CDS

/note="Correia element; hmms hit to HMM Correia (1 - 156), score: 283.48"
/label=Correia
133189. .133193
133200. .133745
/gene="NMA1188"
133200. .133745
/gene="NMA1188"
/note="NMA1188, possible amidase, len: 181 aa; shows very weak similarity to TR:P89923 (EMBL:Y11476), ply12, Bacteriophage 12826 N-acetylmuramoyl-L-alanine amidase endolysin (EC 3.5.1.28) (257 aa), fasta scores; E(): 0.16, 26.0% identity in 177 aa overlap and to bacterial amidases (some putative) e.g. TR:O32421 (EMBL:D76414), lytH, Staphylococcus aureus putative N-acetylmuramoyl-L-alanine amidase (291 aa), fasta scores; E(): 0.0093, 25.8% identity in 182 aa overlap. Identical to NMA1303 and highly similar to NMA1865, fasta scores, E(): 0, 96.7% identity in 181 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="CAB84449.1"
/db_xref="GI:7379879"
/db_xref="SPTREMBL:O9JR25"
/translation="MGKTVTLTAGHSNTDPGAVNGSDREADLAQDMRNIVAAILRDY GLTVKDTGKGNMPLRGSDVAIEFHTNAAVSKAATGEALSTVKNRKCQ VLSKAVAKKGTGWLKRGEDGFKPDNAGHSRLAYAQAGGIVPEFFISNDDLALFKIT KWGICRAIDAIAMELGAARV"
133951. .134115
/gene="NMA1189"
133951. .134115
/gene="NMA1189"
/note="NMA1189, len: 54 aa; unknown, identical to NMA1304 and NMA1863"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1189"
/protein_id="CAB84450.1"
/db_xref="GI:7379880"
/db_xref="SPTREMBL:O9JRV3"
/translation="MGQVAFYERKMWGLWSAKSRSRASEQADLAPEFAGEGELANYQEMI KRHLQTKSVE"
134118. .134354
/gene="NMA1190"
134118. .134354
/gene="NMA1190"
/note="NMA1190, len: 78 aa; unknown, identical to NMA1305 and NMA1862"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1190"
/protein_id="CAB84451.1"
/db_xref="GI:7379881"
/db_xref="SPTREMBL:O9JQP9"
/translation="MRILDFIKNPATGNVSHSKLWNVACAGTVKFVMLPDPSAEIIV AVLGIYGVYAVARSLYSVKRQEVENESREPAGE"
134314. .134319
/gene="NMA1190"
134326. .134745
/gene="NMA1191"
134326. .134745
/gene="NMA1191"
/note="NMA1191, len: 139 aa; unknown, highly similar to NMA1306, fasta scores; E(): 0, 97.1% identity in 139 aa overlap and NMA1861, fasta scores; E(): 0, 97.1% identity in 139 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1191"
/protein_id="CAB84452.1"
/db_xref="GI:7379882"
/db_xref="SPTREMBL:O9JQU5"

repeat_unit


```
HLSPLLEGLPKGTYYADKGYDSENRQHLKEHRLQDGMRRKACNRRLPSEAOQTKNR
A"
complement(138088. .138103)
/notes="inverted repeat at end of IS1106A3"
complement(138173. .139951)
/genes="NMA1200"
complement(138173. .139951)
/genes="NMA1200"
/notes="NMA1200, probable surface fibril protein, len: 529
aa; similar to the N-terminal regions of PR:Q48152
(EMBL:U38617), hia, Haemophilus influenzae adhesin (1098
aa), fasta scores: E(): 0, 44.0% identity in 627 aa
overlap and TR:P71401 (EMBL:U41852), hsf, Haemophilus
influenzae type B surface fibril protein (2353 aa), fasta
scores: E(): 0, 47.8% identity in 584 aa overlap"
/codon_start=1
/transl_table=11
/product="putative surface fibril protein"
/protein_id="CAB84461.1"
/db_xref="GI:7379891"
/db_xref="SPTREMBL:Q9JQW4"
/translation="MNKIYRIIWSALNARVAVSELTRNHTKRASATVKTVATLILF
ATVQANATDEDEELESVQSVGSIQASMEGSELETISLSMTNDSREFVDPYILV
TLKAGNLKIKONTNENTNASSFTYSLKKDLTGLINVETEKLSFGANGKKVNIISDTK
GLNFAKETAGTNGDITVHLNGIGSTLDTLAGSSASHVDAGNQSTHYTRAASIKDVLN
AGNIIKGVGTSTTGOSENVDFVRYTVEFLSADTKTIVNVESKDNKRTEVKIGA
KTSVIREKDKLVTKGKGKNGSSTDEGELVYATKEVIDAVNKAGWRMKTITANGOTG
QADKFETVSGNVTFSKGTGTATYKDDQGNITVYDVNVGDALNVQLNGSNVL
DSKAVAGSGKISVSPSKGMDETVINAGNNIEISRNGKINIDIIATSMAPQFSNL
SLIGADADAPTLSDVDDGALNVGSKDANKPVRITNVPAGKEGDTNVAQLKGVAQLN
NRIDNVDGNARAGIAQAIATAGLVQAYLPKGSMAITGGGTYRGEAGYATGYSSISDGG
NWIIKGTASGNSRGHFGASVGYQW"
repeat_unit
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=RS15
140096. .140156
/notes="91% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTATCCCGATAAGTTTCCGCGACCGACAGAC
CTAG"
repeat_unit
/label=RS15
140157. .140176
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(140177. .140233)
/notes="99% identity to consensus
CCAGCAACACACCGCAGGAGCTTTATCGGAAAAACAGAAACCCCGCCGCGTC"
/label=RS42
140234. .140253
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(140254. .140350)
/notes="100% identity to consensus
GACGAGATTTAGATTATGGGAATTTATCAGGAATGATTGAATCCATAGAAAAACCA
CAGGAATCTATCAGAAAAACAGAAACCCCGCCGCGTC"
/label=RS53
140351. .140370
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
140371. .140431
/notes="91% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTATCCCGATAAGTTTCCGCGACCGACAGAC
CTAG"
/label=RS15
140432. .140451
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(140451. .140507)
/notes="90% identity to consensus
CCAGCAACACACCGCAGGAGCTTTATCGGAAAAACAGAAACCCCGCCGCGTC"
/label=RS42
140508. .140527
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
repeat_unit

/label=DRS3
complement(140528. .140624)
/notes="100% identity to consensus
GACGAGATTTAGATTATGGGAATTTATCAGGAATGATTGAATCCATAGAAAAACCA
CAGGAATCTATCAGAAAAACAGAAACCCCGCCGCGTC"
/label=RS53
140625. .140644
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
140645. .140705
/notes="91% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTATCCCGATAAGTTTCCGCGACCGACAGAC
CTAG"
/label=RS15
140706. .140725
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(140725. .140781)
/notes="86% identity to consensus
CCAGCAACACACCGCAGGAGCTTTATCGGAAAAACAGAAACCCCGCCGCGTC"
/label=RS42
140782. .140801
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(140801. .140897)
/notes="95% identity to consensus
GACGAGATTTAGATTATGGGAATTTATCAGGAATGATTGAATCCATAGAAAAACCA
CAGGAATCTATCAGAAAAACAGAAACCCCGCCGCGTC"
/label=RS53
140898. .140917
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
140918. .140978
/notes="92% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTATCCCGATAAGTTTCCGCGACCGACAGAC
CTAG"
repeat_unit
140979. .140998
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(140999. .141095)
/notes="98% identity to consensus
GACGAGATTTAGATTATGGGAATTTATCAGGAATGATTGAATCCATAGAAAAACCA
CAGGAATCTATCAGAAAAACAGAAACCCCGCCGCGTC"
/label=RS53
141096. .141115
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
141116. .141176
/notes="92% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTATCCCGATAAGTTTCCGCGACCGACAGAC
CTAG"
/label=RS15
141177. .141196
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
141198. .141253
/notes="81% identity to consensus
GACGTCGGGGGTTCCTGTGTTTTTCCGATAGATTCCTCGGTGTTTTCGGTTACTGG"
/label=RS72
141254. .141273
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(141274. .141330)
/notes="98% identity to consensus
CCAGCAACACACCGCAGGAGCTTTATCGGAAAAACAGAAACCCCGCCGCGTC"
/label=RS42
141331. .141350
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
141488. .141498
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
/label=DRS3
141499. .141519
/notes=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
repeat_unit
```

```

/note="Correia element; hmms hit to HMM Correia (1 - 156), score: 277.40"
/label=Correia
/complement(141821..141991)
/gene="NMA1201"
/complement(141821..141991)
/gene="NMA1201"
/note="NMA1201, probable rubredoxin, len: 56 aa; similar to many e.g. SW:RUBR-ACICA (EMBL:246863) Acinetobacter calcoaceticus rubredoxin involved in the hydrocarbon hydroxylating system (54 aa), fasta scores: E(): 1.6e-13, 64.2% identity in 53 aa overlap. Contains pfam match to entry PF00301 rubredoxin, Rubredoxins and P500202 Rubredoxin signature"
/codon_start=1
/transl_table=11
/product="putative rubredoxin"
/protein_id="CAB84462.1"
/db_xref="GI:7379892"
/db_xref="SPTREMBL:Q9JQ3"
/translation="MAQIMGPGCWIIYDEEHGDEHGIAPGTKFEDIPDDWKPECGV GKEDFYLLDRVI"
/complement(141845..141991)
/gene="NMA1201"
/note="Pfam match to entry PF00301 rubredoxin, Rubredoxins, score 97.40, E-value 1.1e-26"
/complement(141863..141895)
/gene="NMA1201"
/note="P500202 Rubredoxin signature"
/complement(141999..142002)
/complement(142012..142091)
/note="Correia element; hmms hit to HMM Correia (76 - 156), score: 105.49"
/label=Correia
/complement(142092..142152)
/note="Correia element; hmms hit to HMM Correia (1 - 61), score: 78.03"
/label=Correia
/label=Correia
142244..142263
/note=">= 90% match to ATTCCCNNNNNNGGGAAT"
/complement(142264..142320)
/label=RS101
/note="RS101"
142321..142340
/note=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=DRS3
/complement(142341..142395)
/note="RS102"
/label=RS102
142396..142415
/note=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=DRS3
/complement(142449..143540)
/gene="NMA1202"
/complement(142449..143540)
/gene="NMA1202"
/note="NMA1202, probable acyl-CoA dehydrogenase, len: 363 aa; shows weak similarity to many e.g. SW:IVD_HUMAN (EMBL:M34192), 1vd, Homo sapiens isovaleryl-CoA dehydrogenase precursor (EC 1.3.99.10) (423 aa), fasta scores: E(): 2.7e-10, 24.0% identity in 358 aa overlap. Contains pfam match to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="putative acyl-CoA dehydrogenase"
/protein_id="CAB84463.1"
/db_xref="GI:7379893"
/db_xref="SPTREMBL:Q9JUQ0"
/translation="MNAQPLIANAEFVKTKLP1VDDIDRKGYYPEAFPMRELGAIIG FVAGTEAGGNGLGATQIAVLREITGECGATSFSAQCAACAWLHOTPNRAVKDK YLADILQKVLAGTGMNTVKHLADIEKHNLQARVEGGYTYNGALPWYINIGEDHIW

```

```

ANTAQIGGVYMFITGCGMEGVSLQNCPEFCALEGTTRTFSLNFKDVFIPDEDIIAAPE
QFADYIQSIRKAGFILLQIGTICAGVIGDSLGIRLVNNAEVNRYLDGQYSLKTRLD
GAWAETERLAGWSGTPDNLTTLKREAAVLTLAAQSAALHSGAGKGYLMRSPAQK
RVREAFVAIVTPAIKHLRKEIAIEAAK"
/complement(143184..143513)
/gene="NMA1202"
/note="Pfam match to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase, score 21.30, E-value 2e-05"
/complement(143547..143550)
/gene="NMA1203"
/complement(143675..144223)
/gene="NMA1203"
/note="NMA1203, len: 182 aa; shows weak similarity to TR:Q53427 (EMBL:S71704), mip, Legionella pneumophila macrophage infectivity potentiator protein (184 aa), fasta scores: E(): 2.9e-08, 26.9% identity in 175 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1203"
/protein_id="CAB84464.1"
/db_xref="GI:7379894"
/db_xref="SPTREMBL:Q9JUP9"
/translation="MARLTVHTLETAPKARVEAVLQNNQFIPNLIGVLSNAPKAL AFYOEVGKLNANSLTAGEVEVIOIITAARTNQCGFCVAGHTKLTALKLLSEQSVKAA RALAAGEFDDAKLGALAAFTQAVMAKKGAVSDEELKAFFDAGYNQQAQVAVVMGVALA TLCNYNNLKGOTEINPELOAYA"
144327..144360
/note="Stem loop containing DNA uptake sequences: gaaat gcgcgtctgaa cattt ttcacgcgc acga"
144332..144341
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/complement(144347..144356)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/complement(144378..144749)
/gene="NMA1204"
/complement(144378..144749)
/gene="NMA1204"
/note="NMA1204, len: 123 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1204"
/protein_id="CAB84465.1"
/db_xref="GI:7379895"
/db_xref="SPTREMBL:Q9JOZ4"
/translation="MKPKIYTHPALPPQAVKQGSWPGLLFCTLWACFKRMNGLGIG LTGAIFLAVFAQLVYGDTPATDSAFNVGLAVSVWFGAKGNSLYARHLLSRGITELP ETVEAANPQALAQYFGRGR"
/complement(144745..144805)
/note="REP 2; hmms hit to HMM REP 2 (69 - 128), score: 48.94"
/complement(144756..144759)
/complement(144803..144845)
/note="Correia element; hmms hit to HMM Correia (114 - 156), score: 47.30"
/label=Correia
/complement(144847..144909)
/note="Correia element; hmms hit to HMM Correia (1 - 62), score: 73.39"
/label=Correia
144981..145022
/note="Stem loop containing DNA uptake sequences: acaat gcgcgtctgaa cggcgctttgcc ttcacagcgc attgc"
144986..144995
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/complement(145008..145017)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
/complement(145044..146735)
/gene="dld"

```

CDS complement(145044. .146735)
/gene="dld"
/EC_number="1.1.1.28"
/note="NMA1205, dld, probable D-lactate dehydrogenase.
len: 563 aa; similar to e.g. SW:DLD_ECOLI (EMBL:M10038),
dld, Escherichia coli D-lactate dehydrogenase (EC
1.1.1.28) (570 aa), fasta scores; E(): 0, 60.9% identity
in 565 aa overlap. Contains Pfam match to entry PF01565
FAD_binding_4, FAD binding domain"
/codon_start=1
/transl_table=11
/product="putative D-lactate dehydrogenase"
/protein_id="CAB84466.1"
/db_xref="GI:3779896"
/db_xref="SPTREMBL:Q9JUP8"
/translation="MSASOLLRLTQTVGEKYLIITDPAKTEORYRGYRFEGRALAVV
RGTILEMKIILQACVEADVIVITQAAANTGLTGGSTPDGNDYDRDVIYVTRRMNLIQ
TINNEQVCLPGSTINLELLELLKPLGREPHSVSGSCIGASVLGVCVGNNSGALVOR
GDAYTEMALFAQINEGRLELVNHLGIDGDTPEELTNLQGHYONKDIKQDAGKGH
DAYCEHVQVDEPTAARENADPARHYEASGCAGKLMFAVRLDTPPEQKQATVFIIG
TNDINELTDIRRAALGEFESLPVSGEYIHRHAFDIADYVKDQTFYVIKFGTQHPKPL
FDLKAVDREGKASFLPKHFSDKAMQVSKYLPDHLPKSMRDYRDKYBHHLLKITGG
KGVDEARAFLEKYFAHHSAGFECAAEETQAAMLHRFAVASAIRVRAVHDEVEDLV
ALDIALRRDRDFEKLPLSEIDNKIHKLYIGHFMCHVFHQDYIIRKKNDCMVALEHEM
LHLQDRGAQYPAEHNVGHLYEAKPALQKQFYRKLDPTNSFNPVGVRTSKKKNAE"
/complement(146064. .146714)
/gene="dld"
/note="Pfam match to entry PF01565 FAD_binding_4, FAD
binding domain, score 153.40, E-value 3.8e-42"
146821. .146855
/notes="Stem loop containing DNA uptake sequences: aaaa
gcgcgtctgaa cygct ttcacagcgc atcgg"
146826. .146835
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(146841. .146850)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(147217. .147226)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
147260. .151093
/genes="NMA1206"
147260. .151093
/genes="NMA1206"
/note="NMA1206, probable oxidoreductase, len: 1277 aa; the
N-terminal half shows weak similarity to several putative
oxidoreductases e.g. TR:Q9ZLW7 (EMBL:AE001480), glcd,
Helicobacter pylori J99 putative glycolate oxidase (459
aa), fasta scores; E(): 1.5e-09, 24.8% identity in 455 aa
overlap. The C-terminal half shows weak similarity to
putative iron-sulphur proteins e.g. TR:O67558
(EMBL:AE000748), gipc, Aquifex aeolicus oxido/reductase
iron sulfur protein (395 aa), fasta scores; E(): 5.2e-07,
22.8% identity in 465 aa overlap. Contains Pfam matches to
entry PF01565 FAD_binding_4, FAD binding domain and to
entry PF00037 fer4, 4Fe-4S ferredoxins and related
iron-sulfur cluster binding domains. Contains PS00198
4Fe-4S ferredoxins, iron-sulfur binding region signature"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAB84467.1"
/db_xref="GI:3779897"
/db_xref="SPTREMBL:Q9JUP7"
/translation="MTTTPAQRIREIPNYNTSYTDREIVIRLLGDEAWQIQLDLRGO
RTGSRARMLFVGDIWVVRNRYLVLDLLEPKRRRAALVPRMRHLNEIRKRRDND
RQDVLVAAAEKAVRFDSSFDTSQKRIILERSKSKITKPNIMFDGLARVTHVTD
TWRYEYPPVYVNPDPTEAEVAPLVRALIELDLVILPRGGGTGTCGAILDMSAVIN
TEKLDKHRGVFEVELAGLDGKHPILRCGAGVTRRVEETAHQGLFAVDPDTSADSC
VGNVAMAGGRKAVLGTALDNLAYQWNPQGEWLRIERVRHNFPGKIHDEATFVDF
VHTLSDGNTIVKTRLEIPGKFRKVGKDVDTKFLSLGFPQDEGDDGITGVAF
VLHKMPKYTRTVMCFEGTVATATPSIVETIROFLAHDSVRVLAGLEHLDWRVVRVGY

ATKAAGKGRPKVLLADVVSDDEAAVEAAAEHICELARADGEGFTAVSPEARKTFWL
DRSRTAAIKHTNAEKNEDVWIPLERIGEYSDGIERINIELSIQNLKLALEQYL
SGKLPTDKGTOLPTAELLGERGKHALAHVSARKERWLLANLADPLADYKARYGSA
VHAAPKADNESCFTAFDRDFRLVSVKADVMPLSEIFSGKTDTKTIOGLGKTHAKTV
RSRVFALHMHAGDGNVHTINPVNSDDAEMLQTAYSRVERIMKIAISLGGVISEGHI
GITKLEFLEDEEMOPFWDYKNQVDPKHTNRHKLKMGSDURNAYTSFELLGRESLIM
RSGDSTIADSVKDCURCKPCVSTHVRNLANLYSPRNKLILGVLTLTFAFLYEQT
RGVSTIKHFEELMDIGDHTCVHRCVPCVNIIDFGDVTAVRNYLADSGHRRFAPAA
AMGMFPLNATGPKTIKALRAAMIOTGTPAQNFAFKIGKLLPICTKQKAEPRATVGKA
PIKEQVIFINRPLPKSVPAKTPSLSLGIEDGSKIPIIRNPAAPDEAAVFPYPGCGS
ERLFSOIGLAVQAMLMHVGVQTVLPFGYMCCTGYPDAGGNKAKAEEMSTNRVAFHRM
ANTLNLDTIKTVSSCGTCTDQLEKRYFEIFFPGCRIIDHIEYLLERGVKLNKGVKQQ
YYHDPCHTPIKTMNATMASSLQKQVLSDRCCGSGMFAVRPDIATQVAKRQOE
ETEKNLKEPQGEPPVKMLTSCFACLOGLSRYTDDNNMPADYIVIEMAKHILGENWLDE
FYKKNNGPQGEKVL"
147662. .148426
/genes="NMA1206"
/note="Pfam match to entry PF01565 FAD_binding_4, FAD
binding domain, score 73.30, E-value 5.2e-18"
complement(148276. .148285)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(148876. .148885)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(149397. .149406)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
149666. .149752
/genes="NMA1206"
/note="Pfam match to entry PF00037 fer4, 4Fe-4S
ferredoxins and related iron-sulfur cluster binding
domains, score 8.70, E-value 0.28"
149882. .149917
/genes="NMA1206"
/notes="PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature"
151105. .151145
/notes="Stem loop containing DNA uptake sequences: gaaa
gcgcgtgaa cgcgcgaagcc ttcacagcgc attgt"
151110. .151119
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(151131. .151140)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
151170. .152171
/genes="NMA1207"
151170. .152171
/genes="NMA1207"
/note="NMA1207, len: 333 aa; similar to many hypothetical
proteins e.g. TR:O68273 (EMBL:AF026544) Alcalligenes
eutrophus hypothetical protein (322 aa), fasta scores;
E(): 0, 47.3% identity in 315 aa overlap. Also similar to
SW:NIR3_RHOCA (EMBL:X72382), nifr3, Rhodobacter capsulatus
nitrogen regulation protein (not stringently required for
nitrogen fixation) (324 aa), fasta scores; E(): 8.1e-19,
31.5% identity in 324 aa overlap. Similar to NMA1633,
fasta scores; E(): 8.3e-20, 34.0% identity in 244 aa
overlap. Contains Pfam match to entry PF01207 UPF00034,
Uncharacterized protein family UPF0034 and PS01136
Uncharacterized protein family UPF0034 signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1207"
/protein_id="CAB84468.1"
/db_xref="GI:3779898"
/db_xref="SPTREMBL:Q9JUP6"
/translation="MIDGQTEPKOKTRIILAPWQGLVDDVMEDLLTRIGGVDECVSE
VRITHIVSRATWLKAYVEPIANGNKTFSGTCTVOLLSGDADNMAALEAVRFEGAD
KIDLNFGCPATVNRHKGAILLKPELFIHVITKRLRPLAHPIPLAKMRLGYDKS
PALEACATAEGGACGLTHARTKAEGYEPFAQHEWIRKIRDSVDIPFVANGDVSFLO
DYIGIKTISGNSVMLGRGAVIRPDLARQIKQYENGPGVKDFTDAEVSVKRQFFELC

| | | | | | | | |
|--------------|---|--------------|---|------|---|------|---|
| misc_feature | 151215. .152153 /gene="NMA1207" /note="Pfam match to entry PF01207 UPF0034, Uncharacterized protein family UPF0034, score 249.90, E-value 3.5e-71" E-value 3.5e-71" 151479. .151535 /gene="NMA1207" /note="PS01136 Uncharacterized protein family UPF0034 signature" complement(152221. .152403) /gene="NMA1208" complement(152221. .152403) /gene="NMA1208" /note="NMA1208, len: 60 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1208" /protein_id="CAB84471.1" /db_xref="GI:7379899" /db_xref="SPTREMBL:Q9JUP5" /translation="MFVVYGFCAFVVALLSLWNAGAFGMQEDDTQSDYERRLGLG AKLNKNTPKSGSKRO" complement(152408. .152412) 152471. .152626 /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 212.79" /label=Correia complement(153062. .153154) /gene="NMA1209" complement(153062. .153154) /gene="NMA1209" /note="NMA1209, len: 30 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1209" /protein_id="CAB84470.1" /db_xref="GI:7379900" /db_xref="SPTREMBL:Q9JUP4" /translation="MOLFCHAVKYPHDKMQNWDKTCARLHKPV" complement(153117. .153425) /gene="NMA1210" complement(153117. .153425) /gene="NMA1210" /note="NMA1210, len: 102 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1210" /protein_id="CAB84471.1" /db_xref="GI:7379901" /db_xref="SPTREMBL:Q9JUP3" /translation="MRKDRKWTQISAAALYGONKTRSAEAVTRSFKNPAVNGKSKGAG ACYGGHLRAIQGFYRACGLRFTVTVAQIKREHLKKQARLEESDRACFATVLSREISA" 153153. .153192 complement(153477. .153992) /gene="NMA1211" complement(153477. .153992) /gene="NMA1211" /note="NMA1211, len: 171 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1211" /protein_id="CAB84472.1" /db_xref="GI:7379902" /db_xref="SPTREMBL:Q9JUP2" /translation="MVVTPKTHQLQYTESNGIAIKTVHRRTSFDSCRATAGKKSREF VESGANKFTVELPDEDGDKTGVVAESGSGTKIGKARELYAPTCAGKPTADEIA PDKAPHEVTAALSAAGLSVPVRLRPTTPRHSACRRDRGRIARMAGYRHRHTVFKHP QRNAVPRKKPL" complement(154017. .154822) /gene="NMA1212" /note="possible IS element remnant" | misc_feature | 151215. .152153 /gene="NMA1207" /note="Pfam match to entry PF01207 UPF0034, Uncharacterized protein family UPF0034, score 249.90, E-value 3.5e-71" E-value 3.5e-71" 151479. .151535 /gene="NMA1207" /note="PS01136 Uncharacterized protein family UPF0034 signature" complement(152221. .152403) /gene="NMA1208" complement(152221. .152403) /gene="NMA1208" /note="NMA1208, len: 60 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1208" /protein_id="CAB84471.1" /db_xref="GI:7379899" /db_xref="SPTREMBL:Q9JUP5" /translation="MFVVYGFCAFVVALLSLWNAGAFGMQEDDTQSDYERRLGLG AKLNKNTPKSGSKRO" complement(152408. .152412) 152471. .152626 /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 212.79" /label=Correia complement(153062. .153154) /gene="NMA1209" complement(153062. .153154) /gene="NMA1209" /note="NMA1209, len: 30 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1209" /protein_id="CAB84470.1" /db_xref="GI:7379900" /db_xref="SPTREMBL:Q9JUP4" /translation="MOLFCHAVKYPHDKMQNWDKTCARLHKPV" complement(153117. .153425) /gene="NMA1210" complement(153117. .153425) /gene="NMA1210" /note="NMA1210, len: 102 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1210" /protein_id="CAB84471.1" /db_xref="GI:7379901" /db_xref="SPTREMBL:Q9JUP3" /translation="MRKDRKWTQISAAALYGONKTRSAEAVTRSFKNPAVNGKSKGAG ACYGGHLRAIQGFYRACGLRFTVTVAQIKREHLKKQARLEESDRACFATVLSREISA" 153153. .153192 complement(153477. .153992) /gene="NMA1211" complement(153477. .153992) /gene="NMA1211" /note="NMA1211, len: 171 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1211" /protein_id="CAB84472.1" /db_xref="GI:7379902" /db_xref="SPTREMBL:Q9JUP2" /translation="MVVTPKTHQLQYTESNGIAIKTVHRRTSFDSCRATAGKKSREF VESGANKFTVELPDEDGDKTGVVAESGSGTKIGKARELYAPTCAGKPTADEIA PDKAPHEVTAALSAAGLSVPVRLRPTTPRHSACRRDRGRIARMAGYRHRHTVFKHP QRNAVPRKKPL" complement(154017. .154822) /gene="NMA1212" /note="possible IS element remnant" | gene | complement(154017. .154822) /gene="NMA1212" complement(154017. .154822) /gene="NMA1212" /note="NMA1212, pseudogene, possible transposase, len: 162 aa; similar to parts of many putative transposases e.g. TR:CA84496 (EMBL:AJ239004) Streptococcus pneumoniae putative transposase (107 aa), fasta scores; E(): 6.9e-22, 57.9% identity in 107 aa overlap. Highly similar to NMA1992, fasta scores; E(): 0.99.0% identity in 103 aa overlap and NMA1993, fasta scores; E(): 0.96.3% identity in 162 aa overlap" /codon_start=1 /pseudo /transl_table=11 /product="putative transposase (pseudogene)" complement(154821. .154862) /note="Correia element; hmms hit to HMM Correia (114 - 156), score: 53.74" /label=Correia complement(154864. .154926) /note="Correia element; hmms hit to HMM Correia (1 - 62), score: 78.71" /label=Correia 155009. .155125 /gene="NMA1214" 155009. .155125 /gene="NMA1214" /note="NMA1214, len: 38 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1214" /protein_id="CAB84474.1" /db_xref="GI:7379903" /db_xref="SPTREMBL:Q9JUP1" /translation="MIIHYKSPSLRDLFDATYPRKEAMIIYPAVLFDQPD" 155138. .155311 /gene="NMA1215" 155138. .155311 /gene="NMA1215" /note="NMA1215, len: 57 aa; similar to parts of hypothetical proteins from bacteriophage and bacteria e.g. SW:Y014_Bphp1 (EMBL:U24159) bacteriophage Hpi hypothetical protein (133 aa), fasta scores; E(): 3e-07, 37.5% identity in 56 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1215" /protein_id="CAB84475.1" /db_xref="GI:7379904" /db_xref="SPTREMBL:Q9JUP0" /translation="MIIPDLPCYVPGDVTADALADTKAAAMFHLEGLQENLPIPGA QSIKVRHNDNPDYA" 155314. .155529 /gene="NMA1216" 155314. .155529 /gene="NMA1216" /note="NMA1216, len: 71 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1216" /protein_id="CAB84476.1" /db_xref="GI:7379905" /db_xref="SPTREMBL:Q9JUN9" /translation="MLMTVEADDAALTGQVRFNISWFOHILNRVDYDAAVMKPERE FGSRFGDAALTFSPISFGFFIADAV" 155521. .155530 /note="Core DNA uptake sequence: gccgtctctgaa" /label=DUS 155557. .156654 /note="IS1106A3" /label="IS1106A3" | gene | complement(154017. .154822) /gene="NMA1212" complement(154017. .154822) /gene="NMA1212" /note="NMA1212, pseudogene, possible transposase, len: 162 aa; similar to parts of many putative transposases e.g. TR:CA84496 (EMBL:AJ239004) Streptococcus pneumoniae putative transposase (107 aa), fasta scores; E(): 6.9e-22, 57.9% identity in 107 aa overlap. Highly similar to NMA1992, fasta scores; E(): 0.99.0% identity in 103 aa overlap and NMA1993, fasta scores; E(): 0.96.3% identity in 162 aa overlap" /codon_start=1 /pseudo /transl_table=11 /product="putative transposase (pseudogene)" complement(154821. .154862) /note="Correia element; hmms hit to HMM Correia (114 - 156), score: 53.74" /label=Correia complement(154864. .154926) /note="Correia element; hmms hit to HMM Correia (1 - 62), score: 78.71" /label=Correia 155009. .155125 /gene="NMA1214" 155009. .155125 /gene="NMA1214" /note="NMA1214, len: 38 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1214" /protein_id="CAB84474.1" /db_xref="GI:7379903" /db_x |
|--------------|---|--------------|---|------|---|------|---|

repeat_unit

```

155557. .155577
/notes="inverted repeat at end of IS1106A3"
/label=IR
155621. .156505
/genes="NMA1217"
15621. .156505
/genes="NMA1217"
/notes="NMA1217, probable transposase for IS1106A3, len:
294 aa; highly similar to many e.g. TR-CAB44967
(EMBL:AJ742841) Neisseria meningitidis transposase for
insertion sequence IS1106A3 (335 aa), fasta scores; E():
0, 79.4% identity in 335 aa overlap. The difference in
length with respect to TR-CAB44967 is accounted for by a
deletion from within the C-terminal region of NMA1217"
/codon_start=1
/transl_table=11
/product="putative transposase for IS1106A3"
/protein_id="CAB84477.1"
/db_xref="GI:7379906"
/db_xref="SPTREMBL:Q9JUN8"
/translation="NSTFFRQTAAQMIKAKHIDRFPLKSDQVIDWQPIEQYLNROTR
YLRDHRGPAYPLLGMFKAVLLGOWHSLSDPELHSLITRIDFNLCRDELSIDPYS
TLCVRNRLAQDITSELLELINRLQTEKLGKVKASAAVVDATIIQTAGSKORQAIIE
VNKGQVSGTTPSKDGDARWTKKNGLYRLGYKQKTRTDAEGYIEKLHITPANTHECN
HLLPLEGTEGTTIYADKGYDSAENRTSGRTSVVRRHYARAAYFLGIKVSQAQSHLKA
MCLNLLKAANRLSAPAAA"
/complement(156634. .156654)
/notes="inverted repeat at end of IS1106A3"
/label=IR
misc_feature
complement(156769. .156778)
/notes="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
gene
complement(156770. .156973)
/genes="NMA1218"
CDS
complement(156770. .156973)
/genes="NMA1218"
/notes="NMA1218, len: 67 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1218"
/protein_id="CAB84478.1"
/db_xref="GI:7379907"
/translation="SPTREMBL:Q9JUN7"
/notes="MELGYTPYNLRNCKLIQAEALQIVGVKHYIQVRWEAEPDPTET
RADMPLEKRWQFLDWTEKTNV"
complement(156773. .159764)
/notes="possible partial prophage element pnm4"
/label=pnm4
gene
complement(156985. .157311)
/genes="NMA1219"
CDS
complement(156985. .157311)
/genes="NMA1219"
/notes="NMA1219, len: 108 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1219"
/protein_id="CAB84479.1"
/db_xref="GI:7379908"
/db_xref="SPTREMBL:Q9JUN6"
/translation="MPDHRKREGLYPPNMKNRAALNRYPEWITIGDQGASRFITHCRY
PLRIAKIHRQTDSECPGGHYRHSNGITLYDFIDGKPADEARFAAVLTETCRRAVK
KICETK"
gene
complement(157304. .157567)
/genes="NMA1220"
CDS
complement(157304. .157567)
/genes="NMA1220"
/notes="NMA1220, len: 87 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1220"
/protein_id="CAB84480.1"
/db_xref="GI:7379909"
/db_xref="SPTREMBL:Q9JUN5"

```

gene
CDS

```

/translation="MADILPSDKRRPPKTKGSTMNIIQEMKHSSLRKVDVEIAKII
ADAHKINAESVKIAOESRWYPMAATGLVTAIAAVALIFKFA"
complement(157587. .158162)
/genes="NMA1221"
complement(157587. .158162)
/genes="NMA1221"
/notes="NMA1221, possible phage integrase, len: 191 aa;
similar to the C-terminal half of phage integrases e.g.
SW:INTA_ECOLI (EMBL:U03737), intA, Escherichia coli
prophage CP4-57 integrase (413 aa), fasta scores; E():
3.1e-06, 30.3% identity in 165 aa overlap"
/codon_start=1
/transl_table=11
/product="putative phage integrase"
/protein_id="CAB84481.1"
/db_xref="GI:7379910"
/db_xref="SPTREMBL:Q9JUN4"
/translation="MMGKPKSRSPALLSARWODIDLTLARLWHIIPADMKTKHPFTV
PLSDWAVEILRELHQTGDNHILFGPIRPRKPAPHPDHISIKFAHNAIRRLGYDST
PGKSKHTMHGFRHLFTNISLTAGKDTLTDTDALGHSRAALQRAGRSSLHHYLTADSY
RLQERRELAEWYSRRHRQAYEAAAQNTCNKT"
complement(158471. .159346)
/genes="NMA1222"
complement(158471. .159346)
/genes="NMA1222"
/notes="NMA1222, len: 291 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1222"
/protein_id="CAB84482.1"
/db_xref="GI:7379911"
/db_xref="SPTREMBL:Q9JUN3"
/translation="MNTETIATDKVSNQAAAMRSVLVMBQWNGGETYSERWVERGR
QAVRQTMGFMFELGRALIIILKEHTHEGRMEIVKVSQFGLGIAETSRMSATREFTPQ
MQRAAPLMDLGRKSLLELVEEDVTILVGLAEGEVNGMTFDDVDRMTYRELVALRDE
SRENLAAKDEVMKOTAKIDEAEKAKKQTVVREPREAEDVGSELAMQLTSLSEVGIRS
QVSRKLDLFDQLNAHSEAHGISHQAKMVGTNLQIILDCQLRRESYALPTAPTNDVYPE
WLGGETGEDESGND"
complement(158676. .158685)
/genes="NMA1222"
/notes="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
gene
complement(159354. .159357)
/genes="NMA1223"
complement(159391. .159681)
/genes="NMA1223"
complement(159391. .159681)
/genes="NMA1223"
/notes="NMA1223, len: 96 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1223"
/protein_id="CAB84483.1"
/db_xref="GI:7379912"
/db_xref="SPTREMBL:Q9JUN2"
/translation="MAESKRVQLRLRVFTALDEHPHIIGLSNKOLSVGLGTPSQVSRD
IDDLVASGLVIKLENGYAYGIKTLQIAERFRQOHERLQSKIAEIGKRVDDV"
complement(159688. .159693)
complement(159693. .159764)
/genes="NMA1224"
complement(159693. .159764)
/genes="NMA1224"
/notes="NMA1224, len: 23 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1224"
/protein_id="CAB84484.1"
/db_xref="GI:7379913"
/db_xref="SPTREMBL:Q9JUN1"
/translation="MVGGQKGNVKGHEIAKKLGLK"
159975. .159978
159986. .162315
/notes="possible partial prophage element pnm5"
/label=pnm5

```

RBS
repeat_unit


```

/db_xref="GI:7379921"
/db_xref="SPTREMBL:Q9JUM4"
/translation="MVGGYALIKRGIAAIPOLAEIKKSANQEGGND"
163336..163339
/gene="NMA1232"
163351..163761
/gene="NMA1233"
163351..163761
/gene="NMA1233"
/notes="NMA1233, len: 136 aa; similar to part of TR-Q50995 (EMBL:L36381) Neisseria gonorrhoeae hypothetical protein (172 aa), fasta scores; E(): 9.1e-15, 61.2% identity in 98 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1233"
/protein_id="CAB84493.1"
/db_xref="GI:7379922"
/db_xref="SPTREMBL:Q9JUM3"
/translation="MIEFVRAKRLWAFVLLVMTGCRYAADKAERKOTALIATYR HSSWVAEOYALQKKQDERWYDFSKQGRKPVKKQYPPQTKKAGYLTKKELLA ELACLKAEAAALKLIDALIYGKEVRQKERNSSQG"
163766..164512
/gene="NMA1234"
/notes="possible IS element"
163766..164515
/gene="NMA1234"
163766..164515
/gene="NMA1234"
/notes="NMA1234, possible transposase, len: 249 aa; similar to many e.g. SW:FI50_ECOLI (EMBL:X07037) Escherichia coli putative transposase for insertion sequence IS150 (283 aa), fasta scores; E(): 0, 43.7% identity in 263 aa overlap. Also highly similar to TR-Q50996 (EMBL:L36381) Neisseria gonorrhoeae hypothetical protein (267 aa), fasta scores; E(): 0, 73.4% identity in 267 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="CAB84494.1"
/db_xref="GI:7379923"
/db_xref="SPTREMBL:Q9JUM2"
/translation="MPSVELLVIVGLPRGTFYYQLAVQSAEDKYADLKRHIHDYRR HKGYGRRIAAAIRHAGTPVNHKVSRLMAKTGLKAVIRWRKVSFSGEYGVKIAPI NLRFCFAEPKNEKWTVDTEFNVGEEKIYLSPIIMDLFNGEIVSYRIQRPYFDLAGEI LKGAPEKPRSEKPTLHSDQGWQOMFYSGNLNQDRAMPQTVQIVRQGEATPYWF KFNPLYNDRILKLLGLSLVQYRIQSLKAA"
164330..164392
/gene="NMA1234"
/notes="Correia element; hmmfs hit to HMM Correia (1 - 62), score: 65.73"
164394..164436
/gene="NMA1234"
/notes="Correia element; hmmfs hit to HMM Correia (114 - 156), score: 51.92"
/label="Correia"
/complement(164836..165198)
/notes="tmRNA: see http://sunflower.bio.indiana.edu/william/tmRNA/seqs/Nme.html"
/label="tmRNA"
/complement(165261..165866)
/gene="NMA1237"
/complement(165261..165866)
/notes="NMA1237, len: 201 aa; similar to bacterial hypothetical proteins e.g. SW:YFCN_ECOLI (EMBL:AE000321), yfcN, Escherichia coli hypothetical protein (183 aa), fasta scores; E(): 3.1e-06, 27.5% identity in 178 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1237"

```

```

/protein_id="CAB84495.1"
/db_xref="GI:7379924"
/db_xref="SPTREMBL:Q9JUM1"
/translation="MNTDFQNLKQLGKQAKAEAKKOKQEDDFDSQAVG QVSLKNCQYIAPPDKPIKVRKPNRADEENFYIGVYNDPASPSSKNGQKNDI QRLKNGYYPVTDVLDHGYTOEEAQKVLNEFIATQKRGVCGEIIHGSLGSKGYKPI LKNMTRNMLQMHPDVLAYAEPRGNDGCVRIILLKRKLQQD"
165876..165876
/complement(165873..165876)
/gene="NMA1239"
/notes="NMA1239, pseudogene, probable transposase, len: 1237 bp; highly similar to many e.g. TR:CAB44967 (EMBL:AJ242841) Neisseria meningitidis strain Z2491 putative transposase for IS1106A3 (335 aa), fasta scores; E(): 0, 94.7% identity in 282 aa overlap. The potential coding region contains a short insertion which results in a frameshift near the C-terminus"
/codon_start=1
/pseudo
/transl_table=11
/product="putative transposase (pseudogene)"
/complement(167583..168601)
/gene="NMA1242"
/notes="IS1106A3 remnant"
/label="IS1106A3"
/complement(167583..168601)
/gene="NMA1242"
/complement(167583..168601)
/notes="NMA1242, pseudogene, probable transposase, len: 1019 bp; highly similar to many e.g. TR:CAB44967 (EMBL:AJ242841) Neisseria meningitidis strain Z2491 putative transposase for IS1106A3 (335 aa), fasta scores; E(): 0, 99.0% identity in 96 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative transposase (pseudogene)"
168154..168308
/notes="Correia element; hmmfs hit to HMM Correia (1 - 156), score: 284.48"
/label="Correia"
/complement(168778..168787)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label="DUS"
168877..168912
/notes="Stem loop containing DNA uptake sequences: gaaat gcgcgtctgaa agcgcg ttcagacggc atagc"
168882..168891
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label="DUS"
/complement(168898..168907)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label="DUS"
/complement(168915..169970)
/gene="NMA1243"
/complement(168915..169970)
/gene="NMA1243"
/notes="NMA1243, probable sulphate-binding protein, len: 351 aa; similar to many e.g. SW:SUBI_ECOLI (EMBL:X02519), sbp, Escherichia coli sulfate-binding protein precursor (329 aa), fasta scores; E(): 0, 49.1% identity in 316 aa overlap. Contains Pfam match to entry PF01100 Sulphate_bind, prokaryotic sulphate- and thiosulphate-binding protein and an appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. Contains a probable N-terminal signal

```

```

sequence"
/codon_start=1
/transl_table=11
/product="putative sulphate-binding protein"
/protein_id="CAB84498.1"
/db_xref="GI:7379925"
/db_xref="SPTREMBL:Q9JUM0"
/translation="MKTYAPALTYAALLTACSPAADSNHPSGQNPANTESDGKNTTL
LNASYVARDFFKEYNPLFIKTYOSEHPTSVSIQSHGSSKQALSVAAGLQADVVT
MNOSSDIDLEKKGVLGKQWQALPDHAAPTSTWFLVRKNPNKQIRDNWDLAKDGV
NIVIANPKTSGNGRYAFLGAYGYGLTKTNGNEOAKLVASILKNTVPFENGGRATT
TFTORNIGDVLITFENEANYVKSLTQGOFEIIVPSYTSAESFPVAVANSYVAKKGTQ
KTRAYLEVLWSEPAQELASLYLRPNPEVLARHADFPDLDTFSPEKFGGWNIM
KTYFAGGIFDRLTAOK"
complement(168921..169970)
/genes="NMA1243"
/notes="NMA1243"
/note="Pfam match to entry PF01100 Sulphate_bind,
Prokaryotic sulphate- and thiosulphate-binding protein,
score 485.20, E-value 5.2e-142"
misc_feature
complement(169920..169952)
/genes="NMA1243"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(169978..169981)
gene
complement(170076..170558)
/genes="NMA1244"
complement(170076..170558)
/genes="NMA1244"
complement(170076..170558)
/notes="NMA1244"
/note="NMA1244, len: 160 aa; similar to TR:P72792
(EMBL:D90900) Synecocystis sp. (strain PCC 6803)
hypothetical protein (160 aa), fasta scores; E(): 1.1e-05,
28.8% identity in 160 aa overlap. Shows very weak
similarity (not statistically significant) to eukaryotic
acetyltransferases e.g. SW:ATDA_MOUSE (EMBL:L10244), sat,
Mus musculus diamine acetyltransferase (EC 2.3.1.57) (171
aa), fasta scores; E(): 1, 24.8% identity in 137 aa
overlap. Contains Pfam match to entry PF00583
Acetyltransf, Acetyltransferase (GNAT) family"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1244"
/protein_id="CAB84499.1"
/db_xref="GI:7379926"
/db_xref="SPTREMBL:Q9JUL9"
/translation="WNLTLRPATVQDCKDIFKVHLHSVQVTCILSYNEHAKYWG
LNTESYLPTISDPDKLWAEYKGIQGFIDQEAQLDALYVHPLFHNILGLGTAL
LHQAEITAHKSLGLSLKYASLNSVPFYLLNRYESLGSVQLQDPSIKIKELMRKHL
"
complement(170151..170540)
/genes="NMA1244"
/note="Pfam match to entry PF00583 Acetyltransf,
Acetyltransferase (GNAT) family, score 32.20, E-value
1.2e-05"
misc_feature
complement(170576..170585)
/label=DUS
/note="Core DNA uptake sequence: gccgtctgaa"
gene
complement(170605..171741)
/genes="purk"
complement(170605..171741)
/genes="purk"
/EC_number="4.1.1.21"
/note="NMA1245, purk, probable
phosphoribosylaminoimidazole carboxylase ATPase subunit,
len: 378 aa; previously sequenced from Neisseria
meningitidis Z2491 as TR:CAB44975 (EMBL:AJ242841). Similar
to many e.g. SW:PURK_BACSU (EMBL:J02732), purk, Bacillus
subtilis phosphoribosylaminoimidazole carboxylase ATPase
subunit (EC 4.1.1.21) (379 aa), fasta scores; E(): 0,
44.5% identity in 375 aa overlap"
/codon_start=1
/transl_table=11
/product="putative phosphoribosylaminoimidazole
carboxylase ATPase subunit"

```

```

/protein_id="CAB84500.1"
/db_xref="GI:7379927"
/db_xref="SPTREMBL:Q9XAY8"
/translation="MKNISLPPAMLGILGGQLGRMFTVAAKTMGYKVTVLDPNPNPA
PAAEADRHLICAPFDNQTALAEELACCAAVTTEFENVNADAMRFIAKHTNVSQDCDVA
IAQNRIOEKAWIRKAGLOTAPYQAICKAEIDITESIQFLPICILKTATLIGDQGQIRV
KTYDELKAFAEHRGVDCVLEKMWDLRGEISVIVCRLLNNDNVOTFPAENIHENGILA
YSTVPAFLSADIQQARQMAORLADELNRYVGLAVEMFVVGDDTHELVVNETAPRPHNS
GHTVDACAQADQOQOQVRLMCNLPNPPADTKLLSSCCMANILGDYWOEGGEPDWFPLQS
RPAHLHLGKKTAKHGRKMGHTIILSTKSDTAPQEAKKLHQSL"
complement(171149..171158)
/genes="purk"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(171738..172283)
/genes="NMA1246"
complement(171738..172283)
/genes="NMA1246"
/notes="NMA1246, len: 181 aa; unknown, previously sequenced
from Neisseria meningitidis Z2491 as TR:CAB44974
(EMBL:AJ242841). Lies within a region of unusually low GC
content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1246"
/protein_id="CAB84501.1"
/db_xref="GI:7379928"
/db_xref="SPTREMBL:Q9XAY9"
/translation="MTYERKLYELRQLRPTTKDGLYTDNEKSEILTVRMSGNTENPK
NFSAVAIGINPSKADSDKTLTQLARFLDMYFTNFKMLNIFSSYSTQGTIGRANT
QTDFSKFCGLCEDADMIILAWGTSRAYSDEKRNRIELFLKAERFMKVCISETGSSS
DTRHPSRISYSQVLQVQFEESA"
complement(172280..173755)
/genes="trpe"
complement(172280..173755)
/genes="trpe"
/EC_number="4.1.3.27"
/note="NMA1247, trpe, probable anthranilate synthase
component I, len: 491 aa; previously sequenced from
Neisseria meningitidis Z2491 as TR:CAB44973
(EMBL:AJ242841). Similar to many e.g. SW:TRPE_PSEAE
(EMBL:D12706), trpe, Pseudomonas aeruginosa anthranilate
synthase component I (EC 4.1.3.27) (492 aa), fasta scores;
E(): 0, 58.0% identity in 495 aa overlap. Contains Pfam
match to entry PF00425 chorismate_bind, chorismate binding
enzyme and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative anthranilate synthase component I"
/protein_id="CAB84502.1"
/db_xref="GI:7379929"
/db_xref="SWISS-PROT:Q9XAZ0"
/translation="MISQEQYQAQAGYNRIPVQELLADLDTPLSYLKLANRPYT
YLLESVVGGERGRYSFGLPSHYLKASKGHVDVYQNGEIVEQHDGNPLPIEAFHN
RFKTEIPSLRPTGLGVGYETIYNFHFARHLKNTKADPTGPTDILLMLSOEL
AVIDNLGKILHVYADPSQDYERARERLEDIRTLQKSCAIPLSLGSKRKTEAVSE
FGEEPFCACVKNIKDYIFAGDCMVVPSQMSMEFTDSPALYRALRTNPSPTLYFY
DFGDFHIVGSSPEILVRBERDDVIVRTAGTRKGTTPAEDLANQDILLSDAKEIAEH
VMLIDLGRNDVGRISKTGEVKYTDKMWIEKISHVWHIVSEVGRLLKDGMTNMDILAAT
FYQSGAGVADSDPASEQWQNTQNKARIVIRAAQMVQEGLDK"
complement(172296..172299)
/genes="trpe"
complement(172310..173098)
/genes="trpe"
/note="Pfam match to entry PF00425 chorismate_bind,
chorismate binding enzyme, score 550.20, E-value 1.4e-161"
misc_feature
complement(172617..172626)
/genes="trpe"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(172817..172840)
/genes="trpe"

```


repeat_unit

178049. .178103
/note="92% identity to consensus
GACGGATTGAGATCGCGCATTTATCGGAGCAACAGAAACCGCTCTCCGTC"
/label=RS24
178104. .178123
/note=">= 90% match to ATTCCCNNNNNNGGAAT"
/label=dRS3
complement(178124. .178180)
/note="85% identity to consensus
CCAGACCCCTGACGCGCGGGAATCTATCGAAATGACTGAAACCCGAGATTCTAG"
/label=RS20
178181. .178200
/note=">= 90% match to ATTCCCNNNNNNGGAAT"
/label=dRS3
complement(178200. .178246)
/note="84% identity to consensus
ATTCACACAAAACAGAAACCAACCAACAGAAACCTAAAAATCTCGTC"
/label=RS14d
complement(178405. .178962)
/gene="NMA1252"
/note="IS1106 remnant"
/label=IS1106
complement(178405. .178962)
/gene="NMA1252"
complement(178405. .178962)
/gene="NMA1252"
/note="NMA1252, pseudogene, probable IS1106 transposase
remnant, len: 186 aa; previously sequenced as TR:CAB44967
(EMBL:AJ242841). Highly similar to many e.g. SW:TRA6_NEIME
(EMBL:Z11857) Neisseria meningitidis transposase for
insertion sequence element IS1106 (ORF 1) (288 aa), fasta
scores: E(): 0, 77.7% identity in 184 aa overlap"
/codon_start=1
/pseudo

/transl_table=11

/product="probable IS1106 transposase remnant

(pseudogene)"

complement(178562. .178604)

/gene="NMA1252"

/note="Correia element; hmms hit to HMM Correia (114 -

156), score: 51.92"

/label=Correia

complement(178606. .178668)

/gene="NMA1252"

/note="Correia element; hmms hit to HMM Correia (1 - 62),

score: 78.71"

/label=Correia

179057. .180275

/note="IS1106A3"

/label=IS1106A3

179057. .179075

/note="inverted repeat at end of IS1106A3"

/label=IR

179120. .180127

/gene="NMA1253"

/label=IS1106A3

179120. .180127

/note="NMA1253"

/note="NMA1253, probable transposase for IS1106A3, len:

335 aa; previously sequenced as TR:CAB44967

(EMBL:AJ242841)"

/codon_start=1

/transl_table=11

/product="putative transposase for IS1106A3"

/protein_id="CAB84508.1"

/db_xref="GI:7379934"

/db_xref="SPTREMBL:Q9S361"

/translation="MSTFROTAAQIAKHIDRFPLKLDQVIDWQPIEQYLNROKNR

YLDRHGRPAYPLLSNFRAVLGQWHSLSDEPLEHSLITRIDNLCRDELELIIPDYS

TICRFLNMQDQLTSLSEKLINCOLTEKGLKIEKASAAVVDATIIQTAGSKORQAEI

VDEQOISQOTPKSDARWIKKGLYKGLKRTDADGYYIEKLIITPANAECK

HLSPLLEGPECTVYADKDYSAENRQHLLEHQLLDGIMRKACNRPLSEVQTKNR

YLSKTRYVVEQSGFTLHRKFRYARAAYFGLIKVSAQSHLKAMCLNLLKAANRLSAPAA

A"

repeat_unit

complement(180257. .180275)
/note="inverted repeat at end of IS1106A3"
/label=IR

gene

complement(180852. .182102)

/gene="glyA"

CDS

complement(180852. .182102)

/gene="glyA"

/EC_number="2.1.2.1"

/note="NMA1254, glyA, probable serine

hydroxymethyltransferase, len: 416 aa; previously

sequenced as TR:CAB44965 (EMBL:AJ242841). Similar to many

e.g. SW:GLYA_ECOLI (EMBL:V00283), glyA, Escherichia coli

serine hydroxymethyltransferase (EC 2.1.2.1) (417 aa),

fasta scores; E(): 0, 64.9% identity in 416 aa overlap.

Contains Pfam match to entry PF00464 SHMT, Serine

hydroxymethyltransferase and PS00096 Serine

hydroxymethyltransferase pyridoxal-phosphate attachment

site"

/codon_start=1

/transl_table=11

/product="putative serine hydroxymethyltransferase"

/protein_id="CAB84509.1"

/db_xref="GI:7379935"

/db_xref="SWISS-PROT:Q9XAY7"

/translation="MFSQSVTLAQDPDLAAIAQEDQROODHVELIASENYVSCAVM

EAQVSQTLNKRYAEGYKPGYGGVEYDIVEQLAIDRVKLFQCAQYANVOPHSGSOAN

QAYVASLKPQDITLQSLAHGGHLLTGASVINSIGLYNAVYGLDENVLDYAEVER

LALHEPKMIVAGASVALOIDWAKPREIADKYAGFLYVDMHYAGVLVAGGEPNYP

FCDFVTTHHTKTLRGPGRGVLICDNTHEKALASSIFPSLQGGPLMHVIAKAVAFKE

ALQPEKQYAKQVKINAAAEELVKRLKIVSGRTESHVFLVDLPKMTGKAAEAA

LGRAHTVNNKALINDPDKPFTSGIRIGSAAATTRGFNEADARVLANLADVLSNPE

DEANLAKVRQVITALCNKYPVGA"

complement(180951. .182081)

/gene="glyA"

/note="Pfam match to entry PF00464 SHMT, Serine

hydroxymethyltransferase, score 847.00, E-value 4.5e-254"

complement(181392. .181442)

/gene="glyA"

/note="PS00096 Serine hydroxymethyltransferase

pyridoxal-phosphate attachment site"

complement(182111. .182115)

complement(182583. .184358)

/gene="ggt"

complement(182583. .184358)

/gene="ggt"

/EC_number="2.3.2.2"

/note="NMA1255, ggt, probable

gamma-glutamyltransferase, len: 591 aa; similar to many

e.g. SW:Ggt_BACSU (EMBL:U49356), ggt, Bacillus subtilis

gamma-glutamyltransferase precursor (EC 2.3.2.2) (587

aa), fasta scores; E(): 0, 38.2% identity in 573 aa

overlap. Contains Pfam match to entry PF01019

G.glu.transsept. Gamma-glutamyltransferase and an

appropriately positioned PS00013 Prokaryotic membrane

lipoprotein lipid attachment site. Contains a probable

N-terminal signal sequence"

/codon_start=1

/transl_table=11

/product="putative gamma-glutamyltransferase"

/protein_id="CAB84510.1"

/db_xref="GI:7379936"

/db_xref="SPTREMBL:Q9JUL8"

/translation="MAKTYLLTALIMSWTISGQVFIHANQKVNTHSAVITGADAHTTP

EHATGGLQVIVTADFVWASANPLATQAGYDILKOGSAAADAWAVOTTLSLVEPQS

SGLGGAFVLYWDNTAKTLTFDQRETPMRAPELFDKDGQPLMEAVVCGRSVG

TVAIPKLMETIHQRVGLPWGKLFDPIDRLAKQGFVSPRLAISVQNOQHLLARYPKT

AAYLPNGVPLQAGSLKLEFADSVQAAQKALHTGKQAQNTIVSVVQNAKDPNG

QLSLQDSYQVVERPCVTVRIYECVCGAPSSGCIAGVQILGTFINFSNQVGYD

ASGLRLDGASRLAFADRDVLDGDPFVPIROLISKYLRHSOLLQSDKALPSV

SAGDFTHEWSSQATPELPSHISIVDKAGNVLSMTTSTENAFGSLMANGYLLNNEL

TFSPREP IQGKQVANRVEPKRPSMAPTIVFKAGKPYMAIGSPGSGRIIGYVAKT

VIAHSDNDNDIQDASAPNLLNRFSGYELETGTFTALQWQQAQNLGDKTDVRLNSGV

QAIIIEPSRLVGGADPRREGRVNGD"

```

misc_feature      complement(182598. .184136)
/genes="ggt"
/notes="Pfam match to entry PF01019 G_glu_transpept,
Gamma-glutamyltranspeptidase, score 527.30, E-value
1.1e-154"
misc_feature      complement(184302. .184334)
/genes="ggt"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
RBS               complement(184364. .184368)
gene              complement(184569. .184792)
/genes="NMA1256"
CDS               complement(184569. .184792)
/genes="NMA1256"
/notes="NMA1256, probable pseudogene, len: 224 bp; similar
to many hypothetical proteins, some of which are
glutaredoxin-like e.g. TR:O30824 (EMBL:AF017750)
Haemophilus ducreyi hypothetical protein (107 aa), fasta
scores; E(): 1.7e-10, 73.8% identity in 42 aa overlap. No
remains of a glutaredoxin active site are apparent in this
CDS"
/concat_start=1
/pseudo
/transl_table=11
/product="hypothetical protein NMA1256 (pseudogene)"
repeat_unit       complement(184788. .184942)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 284.48"
/label="Correia
184927. .185154
CDS               /genes="NMA1257"
184927. .185154
/genes="NMA1257"
/notes="NMA1257, len: 75 aa; unknown"
/concat_start=1
/transl_table=11
/product="hypothetical protein NMA1257"
/protein_id="CAB84512.1"
/db_xref="GI:7379937"
/db_xref="SPTREMBL:Q9JUL7"
/translation="MLIHVYVGAENLCVKRDIIVNFINRTAVLKENFAVSKTLTKDN
IFIVINVKIQDFLTVNOERKAICSNQTVI"
RBS               185117. .185121
/genes="NMA1257"
gene              185130. .185348
/genes="NMA1258"
CDS               185130. .185348
/genes="NMA1258"
/notes="NMA1258, len: 72 aa; similar to hypothetical
proteins TR:Q9ZJY4 (EMBL:AE001544) Helicobacter pylori J99
hypothetical protein (76 aa), fasta scores; E(): 1.1e-06,
47.0% identity in 66 aa overlap and TR:O25839
(EMBL:AE000629) Helicobacter pylori 26695 hypothetical
protein (76 aa), fasta scores; E(): 1.3e-06, 45.5%
identity in 66 aa overlap"
/concat_start=1
/transl_table=11
/product="hypothetical protein NMA1258"
/protein_id="CAB84513.1"
/db_xref="GI:7379938"
/db_xref="SPTREMBL:Q9JRB9"
/translation="MFPEYRDLISKLEQNSFARLFDEHNELDKDKITGLVNNPVTSG
ATIDELKKAKLKDLYAILQKAAGK"
stem_loop         185367. .185401
/notes="Stem loop containing DNA uptake sequences: aaaa
gccgtctgaa aacct ttcacacgac attt"
185372. .185381
misc_feature      /notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
misc_feature      complement(185397. .185396)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
gene              complement(185507. .186481)

```

```

CDS               /genes="fbp"
complement(185507. .186481)
/genes="fbp"
/EC_number="3.1.3.11"
/notes="NMA1259, fbp, probable fructose-1,6-bisphosphatase,
len: 324 aa; similar to many e.g. SW:Fl6P_XANFL
(EMBL:X17252), cfxF, Xanthobacter flavus
fructose-1,6-bisphosphatase (FC 3.1.3.11) (364 aa), fasta
scores; E(): 0, 47.8% identity in 322 aa overlap. Contains
Pfam match to entry PF00316 FBPase,
Fructose-1,6-bisphosphatase"
/concat_start=1
/concat_table=11
/product="putative fructose-1,6-bisphosphatase"
/protein_id="CAB84514.1"
/db_xref="GI:7379939"
/db_xref="SPTREMBL:Q9JUL6"
/translation="MDTLRFELPHELOQNLPALGGVLLSVVSACTEINAKVRLGL
AGVLGMAGTGNIOGEDQKLDVIANNIMIDTLKANPAVAGLASEEEDSFVSAGENRY
LVLFDPDLGSSNIDVNIISVGTIFSIILEKPEGALATESFLQTGRQOLAAGVLYGPTQ
LVFTGCHGVMTFLNAENFVLTKEPKVPSTKEFAINMSRRHMLPPVQOYIDELL
AGETGTGKYNMRWVASWVAELHRIILMRGQVFMYPQDKRPSKCKLRLMYEAPMA
LILQAGASASNARQDILGIRPESLHORVAVIMGSSEEDVYLNRLHSK"
/genes="fbp"
misc_feature      complement(185537. .186310)
/notes="Pfam match to entry PF00316 FBPase,
Fructose-1,6-bisphosphatase, score 467.30, E-value
1.2e-136"
misc_feature      186508. .186517
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
misc_feature      complement(186710. .186719)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(186727. .187572)
gene              /genes="NMA1260"
complement(186727. .187572)
/genes="NMA1260"
/notes="NMA1260, len: 281 aa; unknown, similar to
SW:YHIR_ECOLI (EMBL:U00039), yHir, Escherichia coli
hypothetical protein (280 aa), fasta scores; E(): 0, 42.3%
identity in 274 aa overlap and SW:YHIR_HAEIN
(EMBL:M62809), HI0441, Haemophilus influenzae
protein (281 aa), fasta scores; E(): 0, 41.9% identity in
272 aa overlap"
/concat_start=1
/transl_table=11
/product="hypothetical protein NMA1260"
/protein_id="CAB84515.1"
/db_xref="GI:7379940"
/db_xref="SPTREMBL:Q9JUL5"
/translation="MLSYRHAFHAGNHADMKHTLFLVLOYENRKDKPYWYIDTHGG
AGVYNLEGSQAQVETRGQIALLRQONLPALSDFAHQILSPELYCSPWUA
QSTIRVGDKLRLFLHPTDFVHLQNNNGEAGLKGQVLRDGYKGLISLLPPPPRR
AVLIDPPYEKQDYWRVTETLKSALKRFESGCVLIWYPCLSREESRKLPEELKLVDP
NYLHAEHLHVHAPKADGFGMHGSMFVINPPYLLAEQLAANLPALTRLLAQDEGARVLL
DSKIR"
misc_feature      complement(187623. .187632)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
gene              complement(187660. .188262)
/genes="NMA1261"
CDS               complement(187660. .188262)
/genes="NMA1261"
/notes="NMA1261, possible integral membrane protein, len:
200 aa; similar to many bacterial hypothetical proteins
e.g. SW:YGIH_ECOLI (EMBL:U28379), ygiH, Escherichia coli
hypothetical protein (205 aa), fasta scores; E(): 7.8e-26,
43.5% identity in 193 aa overlap. Contains hydrophobic,
possible membrane-spanning regions"
/concat_start=1
/transl_table=11
/product="putative integral membrane protein"

```

```

/protein_id="CAB84516.1"
/db_xref="GI:7379941"
/db_xref="SPTREMBL:Q9JUL4"
/translation="MFNIPAVVSYLIGSLFSFAVIVSKYYGMDPRTYSGNPGCATNV
LRSGKKAALTLIGDAAGLVAIVLLAQEPLGLSDSAIAAVALAALVGHWPVPF
GFGGKGVALGVALSPTTALVCALIVLWAFGPKVSSLAALTATIAAPLAALFF
MPHTSWIFATLATAILLVLRHKSNLNLKIGKESKIGKR"
misc_feature
188318..188674
/gene="folB"
CDS
188318..188674
/gene="folB"
/EC_number="4.1.2.25"
/notes="NMA1262, folB, probable dihydroneopterin aldolase,
len: 118 aa; similar to many e.g. SW:FOLB_BACSU
(EMBL:M34053), folB, Bacillus subtilis dihydroneopterin
aldolase (EC 4.1.2.25) (120 aa), fasta scores; E():
3.5e-07, 35.3% identity in 119 aa overlap and
SW:FOLB_ECOLI (EMBL:U28379), folB, Escherichia coli
probable dihydroneopterin aldolase (122 aa), fasta scores;
E(): 9.2e-15, 40.0% identity in 115 aa overlap"
/codon_start=1
/transl_table=11
/product="putative dihydroneopterin aldolase"
/protein_id="CAB84517.1"
/db_xref="GI:7379942"
/db_xref="SPTREMBL:Q9JUL3"
/translation="MDKIFLHGKADTLIGVYGERERLQTLIVLDLIGVPEKAGSDD
DIANTVHYAEVCTLRRLHKEQDFLLLEALAEYIADVLGYFGAVVVRKIVKPGILE
GREVGVETGERKRED"
RBS
188696..188700
188708..189244
/gene="NMA1263"
CDS
188708..189244
/gene="NMA1263"
/notes="NMA1263, len: 178 aa; unknown, similar to many
bacterial hypothetical proteins e.g. SW:YOKG_BACSU
(EMBL:D84432), ykG, Bacillus subtilis hypothetical
protein (185 aa), fasta scores; E(): 6.9e-21, 37.9%
identity in 177 aa overlap. Contains Pfam match to entry
PF00293 mutt, Bacterial mutt protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1263"
/protein_id="CAB84518.1"
/db_xref="GI:7379943"
/db_xref="SPTREMBL:Q9JUL2"
/translation="MDLREVKLGGTIYEGGFVSISRDKVRPLPNEGORIVIRHPGA
ACVLAVTDEKVLVROWRYAANOATLELPAGKLDVAGEDMAACALRELAETPTTD
SVRLLYSVTAVGFCNEKMYLFEAEVRLGSLTLANDEDELTITVLMSEEVRLQALND
EIKDKGTLIGLYWLMKD"
misc_feature
188873..189010
/gene="NMA1263"
/notes="Pfam match to entry PF00293 mutt, Bacterial mutt
protein, score 22.30, E-value 7.3e-05"
189589..189625
complement(189668..189822)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 271.05"
/label="Correia"
complement(189900..190259)
/gene="NMA1264"
CDS
complement(189900..190259)
/notes="NMA1264, possible integral membrane protein, len:
119 aa; similar
bacterial putative membrane proteins e.g.
SW:CRCB_ECOLI (EMBL:AF000167), crcB, Escherichia coli
protein which may be involved in chromosome folding (127
aa), fasta scores; E(): 4.3e-13, 42.3% identity in 123 aa
overlap. Contains hydrophobic, possible membrane-spanning
regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"

```

```

/protein_id="CAB84519.1"
/db_xref="GI:7379944"
/db_xref="SPTREMBL:Q9JUL1"
/translation="MLSNIIPLSIGAALGATARLLNLAVPASLSPATGNLFANWTGA
FLICIFAEYNHQKLLITGLTGLSGSTYTLGSGFLETVLQSNRPASALANIFLHTA
GSLLLTWLGKIGTAVK"
misc_feature
190276..190285
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(190288..190968)
/gene="NMA1265"
CDS
complement(190288..190968)
/notes="NMA1265, probable integral membrane protein, len:
226 aa; unknown, contains hydrophobic, probable
membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84520.1"
/db_xref="GI:7379945"
/db_xref="SPTREMBL:Q9JUL0"
/translation="MLFOTVWFSDMVLVSIVILILILAASAPSAFSLARYSALPL
CTVIFSAWCLNASAGGQLAQNNYHLLAVNLVAMVGTSAALWLAALLMPLCYLLFA
DSAGAYPPNALVILIPALVNVRLSRMLNRPNIIFITFVNGFLASAGLITGLIVL
TGILDAANAPFSEILMTTALPFIILAWAEAFLSGISTAFVALKPHWINTFDNRYL
KSDRGITWR"
gene
191133..194177
/gene="NMA1266"
CDS
191133..194177
/gene="NMA1266"
/notes="NMA1266, possible cell-division protein, len: 1014
aa; similar to many e.g. SW:FTSK_ECOLI (EMBL:249932),
ftsK, Escherichia coli cell division protein (1329 aa),
fasta scores; E(): 0, 40.0% identity in 1015 aa overlap,
the E.coli protein being approx. 300 aa longer at the
N-terminus. Note that this means most of the predicted
membrane spans in the E.coli homologue are absent from
this protein. Similar over the C-terminal approx. 500 aa
to NMA1527, fasta scores; E(): 0, 57.9% identity in 556 aa
overlap. Contains Pfam match to entry PF01580
FtsK_SpoIIIE, FtsK/SpoIIIE family and PS00017
ATP/GTP-binding site motif A (P-loop). Contains a probable
N-terminal signal sequence and a hydrophobic, probable
membrane-spanning region"
/codon_start=1
/transl_table=11
/product="putative cell-division protein"
/protein_id="CAB84521.1"
/db_xref="GI:7379946"
/db_xref="SPTREMBL:Q9JUK9"
/translation="MFWIVLIVILLALAGLFFVRAQSEREMREYSAMQEKKEKQOA
ELPEIKDGMDFPELALMLFAVKTVYWLFGVVFRCRNYLAHESPEPVPVPASAN
RADVPTASGYSDSNGTBEAEETEEAAEEAADTEDIATVIDNRRIPFDRISAEQ
LMPSELSIPVPVFEITLEETRAALSAALRETKRYIDAFENETAVPKVRSVDT
PMEGIIIGLDDPLQRTYRMFDADKEAFSESADYGFPEYKQPSAFSAVKAENA
RNAPFRRHAGQGQAEAKSPDVSOGSDGTAVRDARRVSVNLKPNKATVSAEA
RTSLRIPESRTVVVKRDVEMPSETENVETETVSSVGYGVPVDETADIHIEPPAPDA
WVVEPEVPKVPMAIDIPPPPVSEIYNRTYEPAGFEQVQVRSRTAETHLADVDLN
GQWQETAAIANDSEGVAERSGGYLSSETAFGSDSQVCPFPVPSRPSRRASDT
EADAFQSEETGAVSEHLPTDILLPLFPNGATQTEBELLENSITIEKLAEFKVK
VKVDSYSGPVITRYEIEPDVGRVNSVLNLEKRLASLGVSARVETILGKTGML
ELPNFKROMIRLSIEFNSEFAESKSLTALGQDITGPVYTDLCAPHLILVAGTTG
SKGSVGVNAMILMLFAKAPEDVRMTIDPKMLEISYESIPIHLAPVYTDKMLAANA
LWNCVNEKRYRLSMFGVNRNLAFGNQIAEAAARGEKIGAPFSLTPDNPELEKLP
FVYVVDFAIDLMTAGKIEELIARLQAAAGLHLILATORPSVDVITGLIKANI
PRIAFQVSSKIDSRTILDMQGAENLLGGQDMLFLPPGTAYFORVHFAPASDEEVRV
VEYLQFQEPDYDDITLSCGMSDDLIGISRGSDGETDPMDYDAVSVYVLKTRASISGV
QALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNA"
191618..191627
/gene="NMA1266"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

```

misc_feature


```

/note="Pfam match to entry PF00682 HMGL-like, HMGL-like,
score 410.80, E-value 1.3e-119"
misc_feature
197311..197361
/gene="leuA"
/note="PS00815 Alpha-isopropylmalate and homocitrate
synthases signature 1"
197872..197913
/gene="leuA"
/note="PS00816 Alpha-isopropylmalate and homocitrate
synthases signature 2"
198838..198879
/note="Stem loop containing DNA uptake sequences: acgat
gccgtctgaa gcataaaaggc ttcacagcgc attgc"
198843..198852
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(198865..198874)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(198885..198927)
/note="Correia element; hnmfs hit to HMM Correia (114 -
156), score: 51.92"
/label=Correia
complement(198929..198991)
/note="Correia element; hnmfs hit to HMM Correia (1 - 62),
score: 78.71"
/label=Correia
199016..199019
199025..199687
/gene="NMA1272"
199025..199687
/gene="NMA1272"
/note="NMA1272", len: 220 aa; unknown, shows weak
similarity to SW:YECA_ECOLI (EMBL:AE000284), yecA,
Escherichia coli hypothetical protein (221 aa), fasta
scores: E(): 5.5e-12, 27.2% identity in 228 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1272"
/protein_id="CAB84525.1"
/db_xref="GI:7379950"
/db_xref="SPTREMBL:Q9JUK5"
/translation="MDSRKFTASRKSELLDAKSEQGNTRCDBVCGFMALLSGP
DKLTPLDWPEVLGDESQFTAAERSEIERVLVAMAMETTAASDKKLPLNLYENEDG
GSDFYWCNAYLGLDVTPTDFEAVDDAFEEFLYPINALGGIYDEEDNGTIRLOFT
EGELAEISELPYALADIYRWQAVINKPQTVRRREGKTRNDPCPGSGRKYKACCG
KN"
199720..199729
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(199770..200621)
/gene="lgt"
complement(199770..200621)
/gene="lgt"
/EC_number="2.4.99.-"
/note="NMA1273, lgt, probable prolipoprotein
diacylglycerol transferase, len: 283 aa; similar to many
e.g. SW:lgt_ECOLI (EMBL:U12289), lgt, Escherichia coli
prolipoprotein diacylglycerol transferase (EC 2.4.99.-)
(291 aa), fasta scores; E(): 0, 51.8% identity in 272 aa
overlap. Contains PS01311 Prolipoprotein diacylglycerol
transferase signature. Contains hydrophobic, probable
membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative prolipoprotein diacylglycerol
transferase"
/protein_id="CAB84526.1"
/db_xref="GI:7379951"
/db_xref="SPTREMBL:Q9JUK4"
/translation="MTHPQDFPVLISIGPLAVRWYALSYILGFLTFLGRRRIAQG
LSVFTKESLDLFLTWGILGVILGVLGYLFYKFSYDLAHLPLDIFKWEGGMSPHGGF
LGVVIAIWLFGKRKHGIGFLKMLMDTVAPLVLPLGLASGRIGNFINGELWGRVTDINAFWA

```

```

MGFPQARYEDLEAAAHNPLNAEWLQQYGMLEPRHPSQLYQFALGICICLPAVVWLFSSKKQ
RPGQVASLFLGGYIGIFRFAEFAEQDDYDGLLTLGLSMGQWLSPMVLGIVGFVR
FGMKQH"
complement(200172..200210)
/gene="lgt"
/note="PS01311 Prolipoprotein diacylglycerol transferase
signature"
200638..200678
/note="Stem loop containing DNA uptake sequences: caaat
gccgtctgaa aggcgatgacg ttcacagcgc atttc"
200643..200652
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(200664..200673)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(200692..201822)
/gene="NMA1274"
complement(200692..201822)
/gene="NMA1274"
/EC_number="2.7.2.8"
/note="NMA1275, argB, probable acetylglutamate kinase,
len: 298 aa; similar to many e.g. SW:ARGB_CORGL
(EMBL:X86157), argB, Corynebacterium glutamicum
acetylglutamate kinase (EC 2.7.2.8) (294 aa), fasta
scores; E(): 0, 45.7% identity in 276 aa overlap"
/codon_start=1
/transl_table=11
/product="putative acetylglutamate kinase"
/protein_id="CAB84528.1"
/db_xref="GI:7379953"
/db_xref="SPTREMBL:Q9JUK2"

```



```

/translation="MESENIISAADKARILAEALPYIRRFSGSVAVIKYGGNAMTEPPA
LKEGFARDVLLKLVGIHPVIVHGSGPQINAMEKVKKGGEFVGQMRVTDKEAMDIVE
MVLGGHVNKEIYMSINTYGGHVGVSRRDDHFTKAKKLLIDTPEONGVDIGQVGTVES
IDTGLVKLIERGCIPIVAPVGVGKEGEAFENIADLVACKLAELNAEKLMMNTNAG
VMDKTGNLLTKLPKPIDELIADGTYLGGMLPKIASAAVAVNGVRATHIIDGRVPA
LLELIFDAGIGSMIUGGEDA"
complement(202987. .203141)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 271.89"
/label=Correia
repeat_unit
203171. .203190
/notes=">= 90% match to ATTCNCCNNNNNNNGGGAAT"
/label=drs3
repeat_unit
complement(203191. .203251)
/notes="85% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTACCCGATAAGTTTCCGACCAGACAGAC
CTAG"
repeat_unit
complement(203242. .203396)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 278.56"
/label=Correia
repeat_unit
complement(203397. .203404)
/notes="partial RS15"
repeat_unit
203405. .203424
/notes=">= 90% match to ATTCNCCNNNNNNNGGGAAT"
/label=drs3
gene
complement(203522. .204028)
/genes="NMA1276"
CDS
complement(203522. .204028)
/genes="NMA1276"
/notes="NMA1276, possible membrane protein, len: 168 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions. Lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84529.1"
/db_xref="GI:7379954"
/db_xref="SPTREMBL:Q9JUK1"
/translation="MLIEOSVANHKVSVTHSAFTMKLLNMTMRPIKFSMVNTLLFI
VICSPFDLLVOLCTILFHSQKIYFITLRFNFVTKSIYMAIIPILYIFTIKKY
YFYSRKVIILSLALSIYFMDYFFSIYSDNLSTETPLHDYIPILINFFSLVSN
FILSFINK"
repeat_unit
204132. .204165
/notes="88% identity to consensus
GACGGCGCATAAGTTCCTCGTGGGACAGACCTAG"
/label=RS64
repeat_unit
204166. .204185
/notes=">= 90% match to ATTCNCCNNNNNNNGGGAAT"
/label=drs3
stem_loop
204281. .204323
/notes="Stem loop containing DNA uptake sequences: tcaat
gcgctctgaa gcgcgaatcggcg ttcagacggc attgc"
misc_feature
204286. .204295
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
misc_feature
complement(204309. .204318)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
gene
complement(204325. .204993)
/genes="NMA1278"
CDS
complement(204325. .204993)
/genes="NMA1278"
/notes="NMA1278, len: 222 aa; unknown, similar to bacterial
hypothetical proteins e.g. TR:O33611 (EMBL:AB004855)
Streptomyces cyaneus protein possibly involved in
morphological differentiation (277 aa), fasta scores; E():
1.5e-15, 29.1% identity in 227 aa overlap and TR:O69629
(EMBL:AL022121) Mycobacterium tuberculosis hypothetical
protein (287 aa), fasta scores; E(): 2.1e-15, 30.2%
identity in 222 aa overlap"
/codon_start=1
/transl_table=11

```

```

/product="hypothetical protein NMA1278"
/protein_id="CAB84530.1"
/db_xref="GI:7379955"
/db_xref="SPTREMBL:Q9JUK0"
/translation="MKNLAIFDLNLTINTDSHAMPQYLIKKGLVDAAATEAQNKKF
YRDYQNGCIDIDAFKFLHPLAPLARYKEELAERHREFMAEYIIPHISPMQMLVQSHQ
MAGDETIVSSTNEFIITPVCRLGTITNIITGQLSGSGRGYTGNTIGTPSLKEGKIT
RLNQLAERGETLESYGKTYFYSDSKNDLPLRLVDEPVAVNPDAELEKEKEKGWPF
LNPK"
complement(204990. .205721)
/genes="NMA1279"
CDS
complement(204990. .205721)
/genes="NMA1279"
/notes="NMA1279, len: 243 aa; similar to SW:YFGE_ECOLI
(EMBL:AE000336), yfgE, Escherichia coli, hypothetical
protein (248 aa), fasta scores; E(): 3.2e-19, 31.3%
identity in 230 aa overlap and SW:YFGE_HAEIN
(EMBL:U32802), Hll225.1, Haemophilus influenzae
hypothetical protein (231 aa), fasta scores; E(): 5.8e-17,
30.1% identity in 226 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop), which may be
fortuitous"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1279"
/protein_id="CAB84531.1"
/db_xref="GI:7379956"
/db_xref="SPTREMBL:Q9JUK9"
/translation="MIKHAKTINYKPYHPSECTVNLQIFDAAHDYPSDFKFLGTE
NAAFLVYLRHKHQFIYVMGEGAGKSHLLQAWAQAELAGKAAEYIDAASMPLTDA
FAEYLAIVDQVEKLGNEQALLFIFNFRNSKGFLLLGSEYTPQOLVIREDLTRM
AYCLVVEVPLTDQEKRIDALASMAAAQVTVDSEIFEYLLKHWRDMSLMMMLDTLD
NYAVTMKRTIPLRLQLKQEQEQ"
complement(205000. .205003)
/genes="NMA1279"
misc_feature
complement(205509. .205532)
/genes="NMA1279"
stem_loop
205727. .205762
/notes="Stem loop containing DNA uptake sequences: acgat
gcgcgtctgaa aacggc ttcagacggc atcgt"
misc_feature
205732. .205741
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
misc_feature
complement(205748. .205757)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
gene
complement(205766. .206314)
/genes="NMA1280"
CDS
complement(205766. .206314)
/genes="NMA1280"
/notes="NMA1280, pseudogene, ABC-transporter ATP-binding
protein, len: 549 bp; similar to the C-terminal half of
many putative ABC-transporter ATP-binding proteins e.g.
TR:O54061 (EMBL:AJ225561), exsE, Sinorhizobium meliloti
putative ABC-transporter ATP-binding protein (606 aa),
fasta scores; E(): 6.4e-14, 34.9% identity in 172 aa
overlap and SW:YE67_HAEIN (EMBL:U32825), Hll467,
Haemophilus influenzae putative ABC-transporter
ATP-binding protein (589 aa), fasta scores; E(): 0, 55.3%
identity in 179 aa overlap. Note that the N-terminal half
of this protein appears to be NMA1331, beyond the
intervening prophage. Contains Pfam match to entry PF00005
ABC_tran, ABC transporter, PS00211 ABC transporters family
signature and PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/pseudo
/transl_table=11
/product="ABC-transporter ATP-binding protein
(pseudogene)"
complement(205781. .206284)
/genes="NMA1280"
misc_feature

```

```

/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 104.10, E-value 2.8e-27"
/complement(205952). .205996)
/genes="NMA1280"
/note="PS00211 ABC transporters family signature"
/complement(206240). .206263)
/genes="NMA1280"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
206315. .235513
/note="possible prophage pnm2; boundaries defined by
NMA1280 and NMA1331, which appear to be the two halves of
the gene containing the integration site"
/label=pnm2
206543. .206548
206557. .206742
/genes="NMA1281"
206557. .206742
/genes="NMA1281"
/note="NMA1281, len: 61 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1281"
/protein_id="CAB84533.1"
/db_xref="GI:7379957"
/db_xref="SPTREMBL:Q9JUJ8"
/translation="MDLINTPALNTAYDGDVVLGITQIYFRRCVTWIFRRYKQFAR
YTLHTFDKEVIAYTQDI"
/complement(206938). .207819)
/genes="NMA1282"
/complement(206938). .207819)
/genes="NMA1282"
/note="NMA1282, possible DNA-binding protein, len: 293 aa;
unknown, contains probable helix-turn-helix at aa 42-63
(Score 1152, +3.11 SD)"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="CAB84534.1"
/db_xref="GI:7379958"
/db_xref="SPTREMBL:Q9JUJ7"
/translation="MPALVTWMIYTKYLVIFCGVFMERIDLALNLVEERSRLGYSRK
NFAEQTCSAESRLYEGSVNIPADFLAAVOLGVDIOVFVFTGHSANLNAVSDDLN
KLDDAIGKSNVNGVINGVNAVIASSGSVVHOINTQNHVIRTRVDSKPGKEHITLE
QASKLOOLVKQAAAEIAKRSKPSIRAIWASLNACHKVPKSLIALSDYDKRAETYLK
KELGRLSNATSKNNDPWRKKKRIATIKLVNKOLEDWLKSYLEKNFVSLESITELSD
EDLQKTYAAVSTKKRKK"
207858. .207862
207867. .208118
/genes="NMA1283"
207867. .208118
/genes="NMA1283"
/note="NMA1283, len: 83 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1283"
/protein_id="CAB84535.1"
/db_xref="GI:7379959"
/db_xref="SPTREMBL:Q9JUJ6"
/translation="MQANEIYKSLKKNISARMLASALGVTNQSVSEVIRNGRSGKRI
AEAIKAIQKDLVEVFPHYKKTDCRDKKIAELKEMLGVC"
208107. .208110
208110. .210093
/genes="NMA1283"
208120. .210093
208120. .210093
/genes="NMA1284"
/note="NMA1284, possible phage transposase, len: 657 aa;
similar to TR:Q38013 (EMBL:X87627) Bacteriophage D3112
transposase A {690 aa}, fasta scores; E(): 0, 34.0%
identity in 623 aa overlap. Similar to NMA1882, fasta
scores; E(): 0, 41.3% identity in 656 aa overlap. Contains
Pfam match to entry PF00552 integrase, Integrase and
PS00017 ATP/GTP-binding site motif A (P-loop)"

```

```

/codon_start=1
/transl_table=11
/product="putative phage transposase"
/protein_id="CAB84536.1"
/db_xref="GI:7379960"
/db_xref="SPTREMBL:Q9JUJ5"
/translation="MEYMDLKSLLNFGLPNMPRTVIGLRKKAQRDGMQTRRRQKGGG
VEYALPIEIRAAIMKRSDELAEKPKTLPQVRPGTAMSAQALAEAKLLNEKQRSVA
DARCAYVAVLGIKEYEYDCSAAKAAVAOFLGLAESKLDLDAVTLGNLEKANDRSRAKVG
ERTLDGWSAYLVAKENATERVALAPKTTKAVKPIESYGLWLPMEFMQPHNIPSAKLAH
SYREFVQAEANPNVNDPNLSMVRVWDKPLLIQMERGKTKTAAAYKSLLPYVYKRDW
GALKPDNVIGDGHSHFKAQVAHVHGRFPKFEVTIIDGCTRFVVGHSVLAESCVAV
SDALRIGVKHFGPLPIIYSDNGGQGTGKTIDHEITGITSLRIGIAGNPGQGRGI
IERWKDNLJEMARQYETFAAGMDSSTKPLMKYKMFSAFNALEKGDLTETEQOKYLK
KLPSWSRFIADYVKCIDYENNNRPHGCELPRHPDGGHYSPKAYRMRLEBDGDTAPDWLSA
EELATMPQOEVRKVQGWLDLFNNSYFETELAEYHKDEVRSYDLSDASAVNVFDMQ
GKFTKRAQDGNTPFAFTARIDQLAEKRRGKIKRAENAIIKLANAEVNPALAEQAAYV
DELGHLCGNVIEAYVLPKTTGTDQDFVLFEADR"
209185. .209208
/genes="NMA1284"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
209236. .209304
/genes="NMA1284"
/note="Pfam match to entry PF00552 integrase, Integrase,
score 17.10, E-value 0.00078"
210094. .210097
210102. .210278
/genes="NMA1285"
210102. .210278
/genes="NMA1285"
/note="NMA1285, len: 58 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1285"
/protein_id="CAB84537.1"
/db_xref="GI:7379961"
/db_xref="SPTREMBL:Q9JUJ4"
/translation="MMDKQKNAFSTELVEKLKLKRALGRIGRQVRQVQIGVPAERNOA
QISKSGNSFTSDGL"
complement(210263). .210272)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
210296. .210299
210308. .211222
/genes="NMA1286"
210308. .211222
/genes="NMA1286"
/note="NMA1286, possible phage transposase, len: 304 aa;
similar to SW:VPB_BPMU (EMBL:X01149) Bacteriophage Mu DNA
transposition protein B (312 aa), fasta scores; E():
8.3e-12, 26.9% identity in 286 aa overlap and to
SW:VPB_HAEIN (EMBL:U32825), H11481, Haemophilus influenzae
putative cryptic Mu-phage dna transposition protein B (287
aa), fasta scores; E(): 1e-17, 28.7% identity in 296 aa
overlap. Similar to NMA1881, fasta scores; E(): 4e-19,
39.9% identity in 276 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop). Contains a probable
helix-turn-helix motif at aa 22-43 (Score 1644, +4.79 SD)"
/codon_start=1
/transl_table=11
/product="putative phage transposase"
/protein_id="CAB84538.1"
/db_xref="GI:7379962"
/db_xref="SPTREMBL:Q9JUJ3"
/translation="MKQINQALOOKLAEKPEKSGMSQOTLARGIDTSPASVSVNLNCT
YAEKGNVETIEPKIEAFLEMODSKAKRELVLGFVSTKTRRIVEVMRAHSGGETV
VIYQAGLGTQAKVKNYCEKNPRAILIEANPSFALVLMKRLAVAKVSAWGLSNDLIF
ESVSDRLSGRLIVDEANLEIRLRALEIRLHDETGGGLVSGMPRVLRVNGLRKHG
ELVQLYSRVSNALNGELSPDDELFAALPADADEETLLELVKHSNGNTRMRMSKLM
RGAVRTANKNGKMQAGIVKKYSLLIR"
210623. .210646
/genes="NMA1286"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

```

misc_feature

```
misc_feature 211223..211232
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
211299..211302
/translation="MAKTRIKQPAIEAAQDKTEVTAIRQIGDLQREVKFLKLETEAGDK
/genes="NMA1287"
211312..211578
/genes="NMA1288"
211312..211578
/genes="NMA1287"
/notes="NMA1287, len: 88 aa; unknown, similar to part of
NMA1880, fasta scores; E(): 0.0086, 36.5% identity in 63
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1287"
/protein_id="CAB84539.1"
/db_xref="GI:7379963"
/db_xref="SPTREMBL:Q9JUJ2"
/translation="MQVLKKVDMKFEVARFWRCPVGLTVGVWCFVAGMALHSCAKE
PEFVAKEPTKVKEMERQADLEVKVWEAQYMSVEEKMGIYVE"
211571..211690
/genes="NMA1288"
211571..211690
/genes="NMA1288"
/notes="NMA1288, len: 39 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1288"
/protein_id="CAB84540.1"
/db_xref="GI:7379964"
/db_xref="SPTREMBL:Q9JUJ1"
/translation="MSDKPILISPAAKKEALDRAVKREIRAKYGDKAIRGCVK"
complement(211804..211813)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
211826..211831
/translation="MIWSVVGGLAVLLGLGLEMPARIVVVLMIGEGHDGYDDN"
212029..212032
/genes="NMA1290"
212041..212208
/genes="NMA1290"
212041..212208
/genes="NMA1290"
/notes="NMA1290, len: 55 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1290"
/protein_id="CAB84542.1"
/db_xref="GI:7379966"
/db_xref="SPTREMBL:Q9JUJ9"
/translation="MNAKEIAEWIEDRGELIMKKDGEGFVIAARAPDGMWKTAEAT
LAQATLWEEV"
212307..212312
/genes="gam"
212318..212836
/genes="gam"
212318..212836
/genes="gam"
/notes="NMA1291, gam, probable host-nuclease inhibitor
protein, len: 172 aa; similar to several e.g. SW:VGAM_BPMU
(EMBL:X04390), gam, Bacteriophage Mu host-nuclease
inhibitor protein (174 aa), fasta scores; E(): 3.9e-24,
46.8% identity in 173 aa overlap"
/codon_start=1
/transl_table=11
```

```
misc_feature 212275..212755
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
212939..212948
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
212967..212976
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
212996..213211
/genes="NMA1292"
212996..213211
/genes="NMA1292"
/notes="NMA1292, len: 71 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1292"
/protein_id="CAB84544.1"
/db_xref="GI:7379968"
/db_xref="SPTREMBL:Q9JUJ7"
/translation="MNITYLTFDQDSLKALNLLVFTLNHLKVLVMDMDGIEDGLSAV
IETASEKADLSAAFNNAVYEQEEKAA"
complement(213215..213823)
/genes="NMA1293"
complement(213215..213823)
/genes="NMA1293"
/notes="NMA1293, possible phage anti-repressor protein,
len: 202 aa; similar to several e.g. TR:Q38585
(EMBL:X95646), ant, Streptococcus thermophilus
bacteriophage Sfi21 anti-repressor protein (287 aa), fasta
scores; E(): 6.5e-07, 31.3% identity in 134 aa overlap and
TR:AAD25485 (EMBL:AF125520), L0142, bacteriophage 933W
hypothetical protein (209 aa), fasta scores; E(): 2.7e-17,
37.8% identity in 201 aa overlap"
/codon_start=1
/transl_table=11
/product="putative phage anti-repressor protein"
/protein_id="CAB84545.1"
/db_xref="GI:7379969"
/db_xref="SPTREMBL:Q9JUJ6"
/translation="MKDNFTMTQLIPTVSGQLDNQALVDADHLHKLFGVETPFSSK
WIORTEYGTQALDFIGVDKIVRTEAGFFGORDKTVQGYLSLDMAKELCMYERND
KGRQARFYFTEMEKQAKALPDVILYRIDALEDAYFQAAPEMALLRYRSMGLNLTGIG
KLDMNPGAVSYRLKKLNDLGFLEYVFKPKMNVQQQSGLLEG"
complement(213220..213224)
/genes="NMA1293"
213903..213937
complement(213955..214155)
/genes="NMA1294"
complement(213955..214155)
/genes="NMA1294"
/notes="NMA1294, len: 66 aa; unknown, similar to NMA1297,
fasta scores; E(): 1.4e-08, 43.9% identity in 57 aa
overlap and NMA1295, fasta scores; E(): 7.1e-07, 51.1%
identity in 47 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1294"
/protein_id="CAB84546.1"
/db_xref="GI:7379970"
/db_xref="SPTREMBL:Q9JUJ5"
/translation="MDSIKRAQEKYALKRVVTVSFNSEKELDLIKFSENLDPSNWK
SKIRGDMSEKFKKKAKKNIDK"
complement(214159..214164)
/genes="NMA1293"
214236..214245
```

```
misc_feature
/!note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(214264..214273)
/!note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(214356..214541)
/!note="NMA1295"
complement(214356..214541)
/!note="NMA1295", len: 61 aa; unknown, similar to NMA1297,
fasta scores; E(): 2.7e-09, 53.8% identity in 52 aa
overlap and NMA1294, fasta scores; E(): 4.1e-06, 51.1%
identity in 47 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1295"
/protein_id="CAB84547.1"
/db_xref="GI:7379971"
/db_xref="SPTREMBL:Q9JUI4"
/translation="MPDEKTLKAKQAEYQRYQKRLMKTVSFHLEREKELVDFMQKE
IPDFSNVKSVMKEKKR"
complement(214520..214621)
/!note="NMA1296"
complement(214520..214621)
/!note="NMA1296"
len: 33 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1296"
/protein_id="CAB84548.1"
/db_xref="GI:7379972"
/db_xref="SPTREMBL:Q9JUI3"
/translation="MAEHRAFPANLSASHTTTILIKDKCLMKKR"
complement(214548..214551)
/!note="NMA1297"
complement(214814..214981)
/!note="NMA1297"
complement(214814..214981)
/!note="NMA1297", len: 55 aa; unknown, similar to NMA1295,
fasta scores; E(): 1.9e-10, 53.8% identity in 52 aa
overlap and NMA1294, fasta scores; E(): 1.1e-08, 43.9%
identity in 57 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1297"
/protein_id="CAB84549.1"
/db_xref="GI:7379973"
/db_xref="SPTREMBL:Q9JUI2"
/translation="WYDEKLAENRKRYEQKRVIKKVSFNAETEKELELYAQNIDFSQW
VKSIRKEIKK"
complement(214989..214993)
/!note="NMA1298"
complement(215205..215387)
/!note="NMA1298"
complement(215205..215387)
/!note="NMA1298", len: 60 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1298"
/protein_id="CAB84550.1"
/db_xref="GI:7379974"
/db_xref="SPTREMBL:Q9JUI1"
/translation="MNYKRGHVKSPFPQONQKMPSENSNAVSDGIFCCRASMPCH
LCGVAVIKIWFVSVP"
complement(215267..215276)
/!note="NMA1298"
complement(215267..215276)
/!note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(215294..215303)
/!note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(215493..215496)

gene
/!note="NMA1299"
/!note="NMA1299", len: 63 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1299"
/protein_id="CAB84551.1"
/db_xref="GI:7379975"
/db_xref="SPTREMBL:Q9JUI0"
/translation="MNTVKEYTAVRERLLNADYLEEVKDRKGTGNTASVEFVPPKIG
ARGYKFKVRKTLVAVDL"
complement(215692..216108)
/!note="NMA1300"
complement(215692..216108)
/!note="NMA1300",
possible membrane protein, len: 138 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84552.1"
/db_xref="GI:7379976"
/db_xref="SPTREMBL:Q9JUH9"
/translation="MSGEKSENPADEQQKMLQEFLEKQDIAVKQKELSRKDEIQ
SHERIALATEAQKQGHQGEIFKEVKRRRLNAVIAISFLIAALVLTALWRDKDVA
IEIVKIGGAVALGYFAGINRGKAQTKLEQRREKDD"
complement(215912..215921)
/!note="NMA1300"
complement(215912..215921)
/!note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(216163..216172)
/!note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(216163..216172)
/!note="NMA1301"
complement(216163..216172)
/!note="NMA1301", len: 164 aa; unknown, similar to part of
TR_038494 (EMBL:M64097) bacteriophage Mu protein E16 (195
aa), fasta scores; E(): 2.7e-05, 46.7% identity in 60 aa
overlap. Identical to NMA1186 and similar to NMA1867,
fasta scores; E(): 1.5e-25, 55.5% identity in 137 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1301"
/protein_id="CAB84553.1"
/db_xref="GI:7379977"
/db_xref="SPTREMBL:Q9JS31"
/translation="MVFIGICVNSYCVYEFEMREALIAKTKIAOKELGLDDGTYR
AVLERVTDKSCADMVSELESVVDNRSGFKPKAKGNPHGPHLRRTSAAMLDKV
EALLTVGGKHWNTAHAMARRMFGDKVEYLDLDTQLHLKLVAAALQIAENRKRTEKAGGDDG
VRKS"
complement(216698..217126)
/!note="NMA1302"
complement(216698..217126)
/!note="NMA1302",
and similar to NMA1866, fasta scores; E(): 1.8e-10, 29.7%
identity in 145 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1302"
/protein_id="CAB84554.1"
/db_xref="GI:7379978"
/db_xref="SPTREMBL:Q9JRW0"
/translation="MGFEKVEHLLPDTVLVDIVDIGLAATEQLVKAIGGAREFKGKG
VPTERLAILVEAIGEVKTHELLOVYGGELYVPCGCKALQLLNHRFPYQEFVKLRDID
KKSGLMAMTKLCPKYGISSTGYTIINEMSRPAAQALF"
complement(217251..217405)

repeat_unit
```

```
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 283.48"
/label=Correia
217423. .217427
217434. .217979
/genes="NMA1303"
CDS
217434. .217979
/genes="NMA1303"
/notes="NMA1303, possible amidase, len: 181 aa; shows very
weak similarity to TR:p89923 (EMBL:Y11476), p1y12,
Bacteriophage 12826 N-acetylmuramoyl-L-alanine amidase
endolysin (EC 3.5.1.28) (257 aa), fasta scores; E(): 0.16,
26.0% identity in 177 aa overlap and to bacterial amidases
(some putative) e.g. TR:Q32421 (EMBL:D76414), lytH,
Staphylococcus aureus putative N-acetylmuramoyl-L-alanine
amidase (291 aa), fasta scores; E(): 0.0093, 25.8%
identity in 182 aa overlap. identical to NMA1188 and
highly similar to NMA1864, fasta scores; E(): 0, 96.7%
identity in 181 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="CAB84555.1"
/db_xref="GI:7379979"
/db_xref="SPTREMBL:Q9JRZ5"
/translation="MGKTVTLTAGHSNTDPGAVGSDREADLAQDMRNIVAAAILRDDY
GLTVKTDGTGKGNMPLREAVKLIIGSDVAIEPHTNAAVSKAATGIEALSTVKNKRWCQ
VLSKAVAKTKGWLKRGEDGFKPDNAGQHSRLAYAQAGGIVFPEPFIISNDTDLALPKTT
KNGICRAIDAIAMELGAARY"
218185. .218349
/genes="NMA1304"
218185. .218349
/genes="NMA1304"
/notes="NMA1304, len: 54 aa; unknown, identical to NMA1189
and NMA1863"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1304"
/protein_id="CAB84556.1"
/db_xref="GI:7379980"
/db_xref="SPTREMBL:Q9JRV3"
/translation="MGQVAFYEKMIGLWSAKSREASEQADLAAFEAEGLANYOEML
KRHLQTKSVE"
218352. .218588
/genes="NMA1305"
218352. .218588
/genes="NMA1305"
/notes="NMA1305, len: 78 aa; unknown, identical to NMA1190
and NMA1862"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1305"
/protein_id="CAB84557.1"
/db_xref="GI:7379981"
/db_xref="SPTREMBL:Q9JQ99"
/translation="MRILDFKNPATGNVSHSKLWANVACAAGTVKFMVLPDPSAEIW
AVYLGIVGYAVARSLVSKRQEVENSRETAGE"
218560. .218979
/genes="NMA1306"
218560. .218979
/genes="NMA1306"
/notes="NMA1306, len: 139 aa; unknown, identical to NMA1861
and highly similar to NMA1191, fasta scores; E(): 0, 97.1%
identity in 139 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1306"
/protein_id="CAB84558.1"
/db_xref="GI:7379982"
/db_xref="SPTREMBL:Q9JSL1"
/translation="WNLYKLANNQPIAIIALVGTLLAVSHHOGYKSAFAKQQAVID
KMEKDAQALLLSAQNARELEQARAEEKYEVKAHVGMALAKQAEVSRLLTKENK
ETENVLTDRKNAGGCGIDGFGHHQLYKRALGYGN"
```

```
misc_feature
218840. .218849
/genes="NMA1306"
/notes="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
218999. .219193
/genes="NMA1307"
218999. .219193
/genes="NMA1307"
/notes="NMA1307, len: 64 aa; unknown, highly similar to
NMA1192, fasta scores; E(): 1.9e-11, 82.8% identity in 29
aa overlap and NMA1860, fasta scores; E(): 1.9e-11, 82.8%
identity in 29 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1307"
/protein_id="CAB84559.1"
/db_xref="GI:7379983"
/db_xref="SPTREMBL:Q9JUH8"
/translation="MPTPPAALMVAPVRPNPPKDGKTVTLLEHAAEFGGYVAELENON
QAWRDWVNSQAAVDGSEGAR"
219178. .219182
/genes="NMA1307"
219190. .219402
/genes="NMA1308"
219190. .219402
/genes="NMA1308"
/notes="NMA1308, len: 70 aa; unknown, highly similar to
NMA1195, fasta scores; E(): 3.3e-24, 84.4% identity in 64
aa overlap and NMA1157, fasta scores; E(): 1.7e-11, 47.5%
identity in 61 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1308"
/protein_id="CAB84560.1"
/db_xref="GI:7379984"
/db_xref="SPTREMBL:Q9JUH7"
/translation="MTYRELQVTVACRHADLEGLSLRAREQEPFVHVSDLLDRAG
IDYAVRMKDQFTTCVFESATPLLM"
219399. .219518
/genes="NMA1309"
219399. .219518
/genes="NMA1309"
/notes="NMA1309, len: 39 aa; unknown, similar to NMA1195,
fasta scores; E(): 6.1e-08, 65.7% identity in 35 aa
overlap and NMA1857, fasta scores; E(): 0.036, 42.9%
identity in 35 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1309"
/protein_id="CAB84561.1"
/db_xref="GI:7379985"
/db_xref="SPTREMBL:Q9JUH6"
/translation="MIGILKKYSVPSDQKQVEAASRHPGAYVRIVFGDVPV"
complement(219434. .219443)
/notes="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
219530. .219796
/genes="NMA1310"
219530. .219796
/genes="NMA1310"
/notes="NMA1310, len: 88 aa; unknown, highly similar to
NMA1196, fasta scores; E(): 3.9e-33, 94.3% identity in 88
aa overlap and similar to NMA1156, fasta scores; E():
1.5e-10, 44.0% identity in 84 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1310"
/protein_id="CAB84562.1"
/db_xref="GI:7379986"
/db_xref="SPTREMBL:Q9JUH5"
/translation="MDFEFGFTLWPIATAAFWFWNGISGRLKEADKRIDDLKEELH
AVKLSYHTKODAKADNTNIAAALERLENLEKVNKLDRRKADKS"
219609. .219618
misc_feature
```

```

/misc_feature
/feature="NMA1310"
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
219793..219990
/feature="NMA1311"
219793..219990
/feature="NMA1311"
/feature="NMA1311", len: 65 aa; unknown, similar to NMA1197,
fasta scores: E(): 1.1e-14, 56.3% identity in 64 aa
overlap and NMA1855, fasta scores: E(): 1.2e-10, 47.0%
identity in 66 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1311"
/protein_id="CAB84563.1"
/db_xref="GI:7379987"
/db_xref="SPTREMBL:Q9JUH4"
/transl_table="MSDPLEVLARIEAKODDMLANQARMDEELQITKKCKKSAAYV
GLGGVIVTTGWELLRAKFGG"
/feature="NMA1312"
219992..220498
/feature="NMA1312"
219992..220498
/feature="NMA1312"
/feature="NMA1312", possible DNA-binding protein, len: 168 aa;
unknown, similar to NMA1198, fasta scores: E(): 0. 73.7%
identity in 167 aa overlap and NMA1854, fasta scores: E():
0. 56.3% identity in 167 aa overlap. Contains probable
helix-turn-helix motif at aa 21-42 (Score 1051, +2.77 SD)
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="CAB84564.1"
/db_xref="GI:7379988"
/db_xref="SPTREMBL:Q9JUH3"
/transl_table="MAHPKEIREKRLRYVSDQTLIEIAMMCEIPTATARSWKRAAK
ESSDWDKVRAYATLGGGIEDLSRLLAGFLVQYQSTMTMLQDTSIEGLMPSERAKL
LTSLSDAFTKTVAAAKVMPETSKLATAIEVLELFGVEVRYPPQHLQAYVELVEPLG
VIEKKYR"
220519..222141
/feature="NMA1313"
220519..222141
/feature="NMA1313"
/feature="NMA1313", len: 540 aa; unknown, highly similar to
NMA1852, fasta scores: E(): 0. 86.5% identity in 539 aa
overlap. Contains PS00017 ATP/GTP-binding site motif A
(P-loop)
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1313"
/protein_id="CAB84565.1"
/db_xref="GI:7379989"
/db_xref="SPTREMBL:Q9JUH2"
/transl_table="MKSKEFLKSLAEYAQLRQIIIEAEVDGFDSPKATAARRAKVFD
PVGYEYFVNTFYPHYIRISPEKSELHAFILRPEIIRSPKGENEAVGAPRGEGSTQ
VTQLFTLVITVQKHAYIVMDSIDQAPMLEAKAELEFNPRKLTDPPEVCGGGRV
WQAGITVANDVKVQVAGSGKRLGRGPPRPDLTLDIDENDQVRNPEQRDKLNA
WLTKVPLPGVGQKDIYIIGTILHDSVLNRTLNPNFPHGIEKFKAMKRPDRMDLW
DRWELFRDGVTAEEAFYLANKEEMERGAQTSWAARGVLALMKIRARDGHATFDSEY
QNDPVSGEDAPFAKSMKFWNDLPSDLVYFGALDPSLGRAGASRDPISAIVIGYQSRMG
RLYVVEAQIKRLPLDIIEDVLCILHRQYRCKLWFVETVQFQEFKDELVKRSARGIP
VPARAVKPSDKLRIETLOPHMANGILLNESPQTLIQFHFHFPKADHDDGPDVAVHM
LWSGAVANCVLEWQSPDTNDFDDEIKSWSR"
220795..220818
/misc_feature
/feature="NMA1313"
/feature="PS00017 ATP/GTP-binding site motif A (P-loop)"
222131..222134
/feature="NMA1313"
222141..223709
/feature="NMA1314"
222141..223709
/feature="NMA1314"
/feature="NMA1314", len: 522 aa; unknown, similar to
SW:YF01_HAEIN (EMBL:U32826), H11501, Haemophilus

```

```

influenzae hypothetical protein (520 aa), fasta scores:
E(): 0, 40.2% identity in 518 aa overlap and contains a
region similar to SW:YF71_HAEIN (EMBL:U32831), H11571,
Haemophilus influenzae hypothetical protein (79 aa), fasta
scores: E(): 7.5e-14, 63.2% identity in 76 aa overlap.
Highly similar to NMA1851, fasta scores: E(): 0, 49.0%
identity in 510 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1314"
/protein_id="CAB84566.1"
/db_xref="GI:7379990"
/db_xref="SPTREMBL:Q9JUH1"
/transl_table="MAKNKTKIQKPEAALQTDVAQITATGRVIAEHPSNFIPTQKM
RALFEDAESGDIRAQHELFADIEERDSIAANMGTRKRALLTLNWRVAPPRNATPEEE
KLSQAYEMMDSLPTLEDLIMDLMDAVGHGFSALEVEWFSGLYLPNFTHRPQSWF
KWQDNGLLRTRENPEGEALWPLGVVHTQKSRVQOARNGLPRTISLWYMFKHAY
HDAEAFLELYGMPIRIGKYGATKKEKNILLRAVAEIGHNAGIMPEGEIEIHNAA
NGWTSAGNPFLQADWCEKSAARLILGQTLTSGADGKSNALGNHNEIRRDLLVSD
AKOVAQTITISQIIGPFIQINYPHADPNRPKFEFTREPKDIAVFADAIFKLDVGVQ
IPESWVRDKLIVDQGEAVLVQVDPNPVRTALAALSAHTVPSKATGRHQEILDG
ALDDALVPEPDFNSQLNPMVRQVAALNACNSYEADAALNALYNPLDNALRTYMQQA
LFTSDILQGDHARA"
complement(222566..222575)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
223696..224991
/feature="NMA1315"
223696..224991
/feature="NMA1315"
/feature="NMA1315", len: 431 aa; unknown, contains regions
similar to SW:YF70_HAEIN (EMBL:U32831), H11570,
Haemophilus influenzae hypothetical protein (169 aa),
fasta scores: E(): 4.6e-25, 48.7% identity in 154 aa
overlap. SW:YF02_HAEIN (EMBL:U32827), H11502, Haemophilus
influenzae hypothetical protein (414 aa), fasta scores:
E(): 3.8e-22, 35.4% identity in 237 aa overlap and
TR:Q46548 (EMBL:U20249) Bacteroides nodosus plasmid
PJ1787 ORF240' (fragment) (240 aa), fasta scores: E():
4.8e-12, 30.4% identity in 194 aa overlap. Similar to
NMA1850, fasta scores: E(): 0, 33.1% identity in 378 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1315"
/protein_id="CAB84567.1"
/db_xref="GI:7379991"
/db_xref="SPTREMBL:Q9JUH0"
/transl_table="MPAPDLGFALSPLPPKAIWLESKKVTAESYRNLTASEIAKYVT
IARMTDLMDLNDIKTSMVESAKSGQSFDDWRKGIINLLSNKWLHPNGHKDIIIDPA
TGEVFGSPRRLEITVTNMTAYNAGQYQYMANIDARPYWYDAYGDSKTRPAHSAI
DGLVYRDDPFWATFYPNGYCRCSVIALSERDVERQRIYQOSTDNLVETHKIYN
KKGDTYLTAYKAPDGLSYTTDRGEDYNAGRNRYRFDLKYDRAHAQFAKEMGAD
FTSTFQLEKEFEYKQRIDIDGKDEQIKIRNALSRQLKFAAGVLSAQFKAEMGLAGM
TRATVLSDDTLVKQVDSREGQNFDSYAFPLDQLQNPHEVIRDNRELIFTARYKGS
ALWMLKYIKGEIYELQSYRISNDKEIAKFMKKVKL"
225101..225517
/feature="NMA1316"
225101..225517
/feature="NMA1316", len: 138 aa; unknown, shows weak
similarity to SW:VPG_BPMU (EMBL:M74911), bacteriophage Mu
G protein (156 aa), fasta scores: E(): 0.0065, 29.9%
identity in 87 aa overlap and SW:VPG2_HAEIN (EMBL:U32831),
H11568, Haemophilus influenzae putative cryptic Mu-phage G
protein 2 (138 aa), fasta scores: E(): 1.3e-22, 47.4%
identity in 137 aa overlap. Similar to NMA1849, E():
1.6e-11, 38.7% identity in 163 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1316"
/protein_id="CAB84568.1"
/db_xref="GI:7379992"

```

```
/db_xref="SPTREMBL:Q9JUG9"
/translation="MIDVKIDNIFVVLKNSQIERLNGNGIENRYLLMRRLESETMHTAVKLN
FRVAGRPKWLGKLYRGDKPLSDSRKLDSPFSLSDNDTALVGTNIVYAAIHNFGGMAG
RNBKVRIPOREFLTLDKDDKQALMDDVDQYFSGSLIP"
225738..225742
225748..226806
/gene="NMA1318"
225748..226806
/gene="NMA1318"
/notes="NMA1318, pseudogene, possible phage protein, len:
1059 bp ; similar to SW:VP1_BPMU (EMBL:M74911)
bacteriophage Mu I protein (fragment) (58 aa), fasta
scores; E(): 0.12, 51.5% identity in 33 aa overlap
SW:VP1_HAEIN (EMBL:U32827), H11504, Haemophilus influenzae
putative cryptic Mu-phage I protein (355 aa), fasta
scores; E(): 5.8e-14, 31.6% identity in 269 aa overlap.
Similar to NMA1848, fasta scores; E(): 2e-19, 39.1%
identity in 261 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative phage protein (pseudogene)"
226910..227380
/gene="NMA1319"
226910..227380
/gene="NMA1319"
/notes="NMA1319, len: 156 aa; unknown, similar to part of
NMA1831, fasta scores; E(): 6.4e-12, 46.3% identity in 95
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1319"
/protein_id="CAB84570.1"
/db_xref="GI:7379993"
/db_xref="SPTREMBL:Q9JUG8"
/translation="MQPVAQVRLSTSSLLSVATALIEAHGEMTAPDILIEVNRMR
RRMQEIALRAVOTAAESGGLTANAVYAYOTAESLRAAAGRNALVAAVINQKP
PLTVRQAPIDGTIHQIAHEFYGDIAEALVRLNPHIHPAFIKRGTLVNSYAK"
227172..227181
/gene="NMA1319"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
227370..228515
/gene="NMA1320"
227370..228515
/gene="NMA1320"
/notes="NMA1320, len: 381 aa; unknown, shows weak
similarity to SW:VPP_BPMU (EMBL:X06796) bacteriophage Mu
tail protein (379 aa), fasta scores; E(): 1.2e-12, 24.9%
identity in 366 aa overlap. Similar to NMA1830, fasta
scores; E(): 0, 44.8% identity in 373 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1320"
/protein_id="CAB84571.1"
/db_xref="GI:7379994"
/db_xref="SPTREMBL:Q9JUG7"
/translation="MQNNSYGAVSVRVGKGEHRHWEYDIDDFLIPADSFDFVIGR
LGEPAIPDLGSGCEVVDGQIVMTGIIQSQRHGKSGGRELSLGRDLAGFLVDCS
APQLNKMVLDLAAKLAAPWQIRAVVLKVENNPALKDIDIEPGTVWQALTHIAN
SVGLHWPFLPDGLTVGGVDYSSPPVATLCWSRTDSRRNIERMDIEWDTNRFSEVTF
LAQSHGRSGDSAKHDLKWYKDPTMTLHRPKTVVVSADNLAALQKQAKQLDWRLE
GFTLTITVGGHTRDGVLPQPGORVHVVDDEHGIDAVFFLMGRREMLSRMDGTQTEUR
LKEDGIWTPDAYPKAEAAARKKRKGVSCHKKGKQKQAEATVFE"
228512..229180
/gene="NMA1321"
228512..229180
/gene="NMA1321"
/notes="NMA1321, len: 222 aa; unknown, shows weak
similarity to SW:VPV_Bpp2 (EMBL:AF063097) bacteriophage P2
baseplate assembly protein V (211 aa), fasta scores; E():
0.85, 24.4% identity in 221 aa overlap and SW:Yf18_HAEIN
(EMBL:U32827), H11518, Haemophilus influenzae hypothetical
```

```
protein (182 aa), fasta scores; E(): 0.0011, 26.8%
identity in 194 aa overlap. Similar to NMA1829, fasta
scores; E(): 1.8e-32, 46.7% identity in 197 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1321"
/protein_id="CAB84572.1"
/db_xref="GI:7379995"
/db_xref="SPTREMBL:Q9JUG9"
/translation="MSLSKLAKKTAQTAKNIGETLRAAFRGKITLVVSSEPIQRVQLS
GLADETLQDLHLQEGYGFASHPPDCSAVPIPLGGNTSHGVIVCSQSGSRIRKLNKPC
ETAIFNHGAKIVIKQIIIEADCQYRVCKQYVNAADAKNAFLVETSAVLTAQ
GOINGNGMAVEGGDGATFGSDVNGQTGGSFNTDGDVVVAGNIISLRQHPTDSIGGKTLP
AEPA"
229253..229262
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
229280..229627
/gene="NMA1322"
229280..229627
/gene="NMA1322"
/notes="NMA1322, len: 115 aa; unknown, shows weak
similarity to SW:YF19_HAEIN (EMBL:U32827), H11519,
Haemophilus influenzae hypothetical protein (135 aa),
fasta scores; E(): 0.031, 36.3% identity in 102 aa
overlap. Similar to NMA1828, fasta scores; E(): 3.4e-20,
51.3% identity in 117 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1322"
/protein_id="CAB84573.1"
/db_xref="GI:7379996"
/db_xref="SPTREMBL:Q9JUG6"
/translation="MDKELNPGTGDYTRVDTLQNAVYILRLMPLGSMWADKTLGSL
LHLQREKDLQVRSLLAEQYADEALQPIVKSGRADKITVRAEQPHDGRLLIHVRDTA
AGGFDYRHEVPVI"
229628..229632
229641..230696
/gene="NMA1323"
229641..230696
/gene="NMA1323"
/notes="NMA1323, len: 351 aa; unknown, similar to bacterial
hypothetical proteins including SW:KKDT_BAC5U
(EMBL:270177), xkdt, phage-like element PbsX protein (348
aa), fasta scores; E(): 1.2e-07, 28.0% identity in 361 aa
overlap and SW:YF20_HAEIN (EMBL:U32827), H11520,
Haemophilus influenzae hypothetical protein (355 aa),
fasta scores; E(): 4.7e-16, 31.5% identity in 321 aa
overlap. Similar at the N-terminus to NMA1827, fasta
scores; E(): 1.5e-18, 41.4% identity in 152 aa overlap and
at the C-terminus to NMA1826, fasta scores; E(): 1e-29,
50.8% identity in 185 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1323"
/protein_id="CAB84574.1"
/db_xref="GI:7379997"
/db_xref="SPTREMBL:Q9JUG5"
/translation="MFETPFEQIRERILRDLTKSLWFPDADISPSDHYVHVASRLASCA
EGQYAHQSWIVRQIFPDPTADREYLERHSMGLRRNPPTTASGTLTVSGTIAQSLSDG
LQVIRGQRFYRTTARAVISGGTAEIPAIADPEGAANVRDGEALMAAGVAGTECR
LTVGGTGRDSDSLARLLEIIRRPAGGNRYDKNMAISVDGVTGAYVYPLRRGLG
TVDITATSDGVPSEETVRRVQAVIDEMRPVAKNALVLPKTVTAVPTVQVKLDGID
LDEAKRIRITAKLEYFDTLIPGDGLTVSQIEAASINVDGVIDLRLTAPTANRAADTVN
RIEWFKAGAINVTEMP"
complement(229937)..229946)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
230681..230684
/gene="NMA1323"
230693..231253
/gene="NMA1324"
230693..231253
```

/gene="NMA1324"
/note="NMA1324, len: 186 aa; unknown, shows weak
similarity to SW:FMFO_ECOLI (EMBL:AE000214), ymfQ,
Escherichia coli hypothetical protein (194 aa), fasta
scores; E(): 0.0001, 28.2% identity in 195 aa overlap.
Similar to NMA1825, fasta scores; E(): 9.6e-24, 42.4%
identity in 184 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1324"
/protein_id="CAB84575.1"
/db_xref="GI:7379998"
/db_xref="SPTREMBL:Q9JUG4"
/translation="MSYQDILRGLLPVSVARNAPRVRAQAEIDGALDAVAESAQSV
ADAVPSAGOMLADWERVLGDTGKRNQRRVLAVAKLNETGGLSIPYFVRLAEAA
GYQIDPEOPFRAGVNRAGDRLAPQETIMWVHVNRGNRRITFRFRAGISAAGDRLT
DYSDAVIESLFNRLKPAHTAIRFYR"
231253. .231256
231264. .233237
/gene="NMA1325"
231264. .233237
/gene="NMA1325"
/note="NMA1325, possible phage tail fibre protein, len:
657 aa; similar in part to SW:VPH_BPH1 (EMBL:U24159)
bacteriophage Hp1 probable tail fibre protein (925 aa),
fasta scores; E(): 0, 48.7% identity in 452 aa overlap.
The N-terminus is similar to that of NMA1824, fasta
scores; E(): 3.9e-09, 33.3% identity in 147 aa overlap"
/codon_start=1
/transl_table=11
/product="putative phage tail fibre protein"
/protein_id="CAB84576.1"
/db_xref="GI:7379999"
/db_xref="SPTREMBL:Q9JUG3"
/translation="MHPETPKDTFHDGCVSELGTILPAWMLNOVSELLAVLTAG
IQPKQPNOLLALNRLVVITANQETGQKTFEAGTQFQSGHLHSANQTNNGGH
AYIGADADNAHIVGDDTLRLHSANNRISYNNHIDIFHKANKPFAEDIEGKPNLISY
YPAITSVYERHQTUUKDTGIYSLPTAVGSSNLPVENTACHIQVITAGTKHGWCRIG
AONTFDLYRAGNSRNLPLDSRTDITAWFSDQIPTGLWAFDDITRTVTKATYPE
LYRLLTGKYGSIONQAEDRFIRNAANGIAGVTKQDEIKRHVKHVFHSWNHPIAA
ALGVDHNERORSALVSTWTDNLNDNGFLTPRPDSKMATGSDENRPKALVLKLIKA
ADTLGEAVFWIKSHGETVNAAGLDAGTLAQNLQDKADRAHTHTAAQIQGLDKEISITAV
AAFTQRTIGVDIVRPDGTMIQTGYSFQSGPIENEVFPVAFADGNVKKCFVSE
RHSGRVTDQRHNLWFIIRAKNHAAAIITNWEGSDMWAIGKAASGNAASSPIGPEI
PETNEPQRESGRSTGPRNRRHRSLLLEALQD"
233387. .233406
/note=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
complement(233407. .233462)
/note="83% identity to consensus
CCAGTCGGTTCAGTTTCAGTCATTTCCGATAAAATTCCTGTTGCTTTTCAFTTCTAG"
/label=RS17
233463. .233482
/note=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
233504. .233523
/note=">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=drs3
complement(233524. .233542)
/note=">= 85% match to ATTCCCNACNACNTTTCGTC"
/label=RS3x
complement(233674. .233976)
/gene="NMA1326"
complement(233674. .233976)
/note="NMA1326, len: 100 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1326"
/protein_id="CAB84577.1"
/db_xref="GI:7380000"
/db_xref="SPTREMBL:Q9JUG2"

/translation="MQROSLTNADGEVRELAAEDFALARPITAEALPEDLAQVLFHQK
QLEKGVNQKRNKTGKTPKQLVITIRLSADVVVEKFRAGGKQWTRINEVLRYVYAQLK"
complement(233948. .234235)
/gene="NMA1327"
complement(233948. .234235)
/gene="NMA1327"
/note="NMA1327, len: 95 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1327"
/protein_id="CAB84578.1"
/db_xref="GI:7380001"
/db_xref="SPTREMBL:Q9JUG1"
/translation="MKIEFSEKQNIERNLPFESVGQIRWTTAVIVPDVCFDYPE
PRYVAAAYLGDTRQLRHIVCFPTPKIDGIRVISFRKANKREVKKYATSIILNKR"
complement(233986. .233990)
/gene="NMA1327"
234302. .234305
234315. .234920
/gene="NMA1328"
234315. .234920
/gene="NMA1328"
/note="NMA1328, len: 201 aa; unknown, C-terminal half is
similar to SW:YF02_HAEIN (EMBL:U32820), HI1402,
Haemophilus influenzae hypothetical protein (68 aa), fasta
scores; E(): 1.3e-08, 56.3% identity in 64 aa overlap.
Shows weak similarity to NMA1823, E(): 3.6e-07, 28.8%
identity in 191 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1328"
/protein_id="CAB84579.1"
/db_xref="GI:7380002"
/db_xref="SPTREMBL:Q9JUG0"
/translation="MTIYFNKGFYDDTLGSTPECAVAVRAEYEAALLAGQGGQIAA
DSGQPVLTTPRSEYHEWDGKWEIGEAARAFQKATATAPRLAAKADLNKSL
AGYQVSDISFYQEKEALAQADNNAPTMLQIAAARGVELDVLIEKVVYKSARLA
VAAGAITGKQQLEDKLTITETAPGLDALEKEIBEWTNLIG"
234886. .234889
/gene="NMA1328"
234899. .235195
/gene="NMA1329"
234899. .235195
/gene="NMA1329"
/note="NMA1329, len: 98 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1329"
/protein_id="CAB84580.1"
/db_xref="GI:7380003"
/db_xref="SPTREMBL:Q9JUF9"
/translation="MDAKHRLKKVYVHLLVAIDQLFNALTGGAADETLSRSTYRGRL
AOKPRTKRWKLYTLINGVCFDRQHCQAYISELKGROHDARFNQSRAGEKGR"
235192. .235401
/gene="NMA1330"
235192. .235401
/gene="NMA1330"
/note="NMA1330, len: 69 aa; unknown, similar to the
C-terminus of SW:YF23_HAEIN (EMBL:U32828), HI1523,
Haemophilus influenzae hypothetical protein (296 aa),
fasta scores; E(): 2.3e-07, 33.9% identity in 59 aa
overlap and to the C-terminus of NMA1821, fasta scores;
E(): 1.7e-07, 39.3% identity in 56 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1330"
/protein_id="CAB84581.1"
/db_xref="GI:7380004"
/db_xref="SPTREMBL:Q9JUF8"
/translation="MNYFGMVFEFLRLMVMVRPPVFFSGTRSELPTYLDLVAELRLTG
WERFAGSOTLTVSSSTTAQATTT"
complement(235511. .236653)
/gene="NMA1331"

CDS

complement(235511.. .236653)
 /gene="NMA1331"
 /note="pseudogene, ABC-transporter ATP-binding protein, len: 1140 bp; similar to the N-terminal half of many putative ABC-transporter ATP-binding proteins e.g. TR:054061 (EMBL:A0225561), exse, Sinorhizobium meliloti putative ABC-transporter ATP-binding protein (606 aa), fasta scores; E(): 4.1e-20, 27.7% identity in 321 aa overlap and SW:YE67_HAEIN (EMBL:U32825), H11467, Haemophilus influenzae putative ABC-transporter ATP-binding protein (589 aa), fasta scores; E(): 0, 49.0% identity in 355 aa overlap Note that the C-terminal half of this protein appears to be NMA1280, beyond the intervening prophage. Contains hydrophobic, possible membrane-spanning regions"
 /codon_start=1
 /pseudo

/transl_table=11
 /product="ABC-transporter ATP-binding protein (pseudogene)"

misc_feature

235583.. .235592

/note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS

misc_feature

236677.. .236686

/note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS

repeat_unit

complement(236760.. .236914)
 /note="Correia element; hmms hit to HMM Correia (1-156), score: 277.83"
 /label=Correia

misc_feature

237058.. .237067

/note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS

gene

complement(237095.. .237742)

/gene="NMA1332"

CDS

complement(237095.. .237742)

/gene="NMA1332"
 /note="NMA1332, possible periplasmic protein, len: 215 aa; unknown, contains a probable N-terminal signal sequence"
 /codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="CAB84583.1"

/db_xref="GI:7380005"

/db_xref="SPTREMBL:Q9JQV3"

/translation="MKPLILGLAVALSACQVOKAPDFYTSFKESKPASTLVVPL NESPDVNTGVLASTAAPLSEAGYVFPAAVVEYTFKQGLTNAADIHAVRPEKLHQ IFGNDALVITVEYTGTSYQLDSYTVTSARAKLVDSRNGELWSGASIREGSNSH SGLIGALVSAYVNOIANSITDRGYQVSKTAAAYNLLSPYSHNGILKGRPFVEQPR"
 complement(237739.. .238110)

/gene="NMA1334"

complement(237739.. .238110)

CDS

/gene="NMA1334"

/note="NMA1334, possible periplasmic protein, len: 123 aa; unknown, contains a probable N-terminal signal sequence"
 /codon_start=1

/transl_table=11

/product="putative periplasmic protein"

/protein_id="CAB84584.1"

/db_xref="GI:7380006"

/db_xref="SPTREMBL:Q9JQJ2"

/translation="MMNPKTILSRLSLCAAVLALTACGGNGOKSLYYGYGPDVYVEGL KNDOTSLCKQTEKMEKYFVEAGNKMNAAPGAHAHLGLLSRSGDKEGAFRQFEKRL LFPESGVFMDLMTGKGGR"
 complement(237749.. .237754)

/gene="NMA1334"

complement(238113.. .238117)

complement(238117.. .238788)

/gene="NMA1335"

CDS

complement(238117.. .238788)

/gene="NMA1335"

/note="NMA1335, possible periplasmic protein, len: 223 aa; unknown, contains a probable N-terminal signal sequence"

/codon_start=1
 /transl_table=11
 /product="putative periplasmic protein"
 /protein_id="CAB84585.1"
 /db_xref="GI:7380007"

/db_xref="SPTREMBL:Q9JRF1"

/translation="MKTSTAVVLAAAVSLTGCAATESRSLEVEKVASYNTOYHGVR TPISVGTDFNRRSFGKIFSDGEDRLGSAKTLVTHLQOTNRFNVLTNINLAKQE TSGISGKHNLCGADYVVTGDDVTEFGRRDVGDHQILGILGRGKSQATAYAKVALNIVNVN TSEIIVYSAQAGEYALSNNRIIGFGTSGYDNLGKVLDLAIREAVNSLVQAVDNGA WQPNR"

complement(238953.. .239723)

/gene="NMA1336"

complement(238953.. .239723)

/gene="NMA1336"

/note="NMA1336, possible oxidoreductase, len: 256 aa;

similar to many from both bacterial and eukaryotic sources e.g. weak similarity to SW:TRN1_DAYST (EMBL:L20473) Datura stramonium tropinone reductase-I (273 aa), fasta scores; E(): 7.1e-11, 30.2% identity in 205 aa overlap. Similar to SW:YCIK_ECOLI (EMBL:AE000225), yciK, Escherichia coli hypothetical oxidoreductase (252 aa), fasta scores; E(): 6.9e-25, 35.2% identity in 233 aa overlap. Contains Pfam match to entry PF00106 adh_short, short chain dehydrogenase"

/codon_start=1

/transl_table=11

/product="putative oxidoreductase"

/protein_id="CAB84586.1"

/db_xref="GI:7380008"

/db_xref="SPTREMBL:Q9JUF7"

/translation="MRQNAKIPTDQHGKGMATLSKDTILITGASQGLGEQVAKAYA AAGATVILVARIHQKLEKVDYDAIVEARHPPEAFIRFDLMSAEKEFNQQAATIAEATQ GKLDGIVHCAGYFALSLPDDFQTVAEWVQYVINTVAPMGLTRALPLLKQSPDASVI FVGESHGTEPKAYWCGFGASKAALNYLCKVAADWERFENGLRANVLVPGPINSQPQRIK SHPGAESKRSYGDVLPFAFVWMASTESGRSGEIVYL"
 complement(239082.. .239654)

misc_feature

/gene="NMA1336"

/note="Pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 130.20, E-value 3.9e-35"
 complement(239755.. .239764)

misc_feature

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS

complement(239756.. .241069)

/gene="NMA1338"

complement(239756.. .241069)

/gene="NMA1338"

/note="NMA1338, possible oxidoreductase, len: 437 aa; shows weak similarity to phytoene dehydrogenases e.g. TR:Q0406 (EMBL:X78815), pds, Narcissus pseudonarcissus phytoene dehydrogenase precursor (570 aa), fasta scores; E(): 1.6e-05, 24.7% identity in 473 aa overlap. Also to

SW:Y4AB_RHISN (EMBL:AE000064), Y4AB, Rhizobium sp. (strain NGR234) hypothetical protein (417 aa), fasta scores; E(): 8.9e-15, 27.3% identity in 444 aa overlap"
 /codon_start=1

/transl_table=11

/product="putative oxidoreductase"

/protein_id="CAB84587.1"

/db_xref="GI:7380009"

/db_xref="SPTREMBL:Q9JUF6"

/translation="MMNTPHRPKPIAVIGAGWAGLSAAVTLARHADVTLFEAGROAGG RARLAGNTDGFGLDNGOIHLLGAVRGVLRMLKTIIGSDPHAAFLRVPLHMHGGLQ FRALPLPAPLHILGGVLLARRVPSAFKALLADMSDLQKSARLGQDPTTVVAOHLQORN VPRAAVQFWQLVWGLNTPLETASLRVLCVSDGLVTKKSGSYDLPKQDLGAI AEPALAEQRLGADLRETRICLNTPDGKVLVNGVEPFDAAVAPATAPYHAALLPEG TPHEVQTAYONLRYHAITVYLYRAEPVRLPAPLTADGLTQVQLCRGLRGLPENVE SAVISVSDRVGAFANRAWADKVHADLKRILPLHLGPEAVRVITEKRATTAAADAPPDL SWLHBRIFPAGDYLHPDYPATLEAAVQSGFSAEACLQSLSDAV"

misc_feature

240219.. .240228

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS

complement(240342.. .240351)

misc_feature


```

/note="Stem loop containing DNA uptake sequences: aacag
gccgtctgaa gccgcacgc ttcacagcgc attgt"
246141. .246170
246141. .246150
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(246161. .246170)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
246236. .246296
/note="Correia element; hmfms hit to HMM Correia (1 - 62),
score: 81.29"
/label=Correia
246298. .246340
/note="Correia element; hmfms hit to HMM Correia (114 -
156). score: 44.42"
/label=Correia
complement(246421. .247008)
/genes="NMA1345"
complement(246421. .247008)
/genes="NMA1345"
/note="NMA1345, len: 195 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1345"
/protein_id="CAB84593.1"
/db_xref="GI:7380015"
/db_xref="SPTREMBL:Q9JUF1"
/translation="MNNFPAKLSKLYVTASDHPDLSSEMEFEEDRLILIRKLYQLLD
GQHILSRVTCLHQNRRRLIALDKAAAGCDAAMLRAQWRYYTOIAERLSSGGGFT
VTAESVSAACPELEGYELVRAAVSFGFTQGQENPAGGQEKTAQVPAASGHSYRK
PIKINIEAVERYPAYRPTAALKVTRRFRFLQNR"
complement(247004. .247139)
/note="REP 2; hmfms hit to HMM REP 2 (1 - 128), score:
129.26"
complement(247013. .247018)
247221. .247418
/genes="NMA1347"
247221. .247418
/genes="NMA1347"
/note="NMA1347, len: 65 aa; unknown, similar to bacterial
hypothetical proteins e.g. TR:O69223 (EMBL:AF010139)
Azotobacter vinelandii hypothetical protein (68 aa), fasta
scores: E(): 2.9e-14, 56.9% identity in 65 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1347"
/protein_id="CAB84594.1"
/db_xref="GI:7380016"
/db_xref="SPTREMBL:Q9JUR02"
/translation="MKWTDTORIAEELYDLHGETIDPRTVRTQLRDLIMALPEFDDO
PARCGERILEAVQQAVIDEAE"
247430. .247439
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(247486. .247887)
/genes="NMA1348"
complement(247486. .247887)
/genes="NMA1348"
/note="NMA1348, len: 133 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YRFH_ECOLI (EMBL:U18997),
yrfH, Escherichia coli hypothetical protein (133 aa),
fasta scores: E(): 1.3e-19, 45.5% identity in 132 aa
overlap. Contains Pfam match to entry PF01479 S4, S4
domain"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1348"
/protein_id="CAB84595.1"
/db_xref="GI:7380017"
/db_xref="SPTREMBL:Q9JUR17"
/translation="MKDKHDSAMRLDKWLWAARFFKTRSLAQKHIELGRVQVNGSKV
KNSKTIIDIGDIIDLTLNSLPYKIKVKGHLNHQRRPASEARLLYEEDAKTATLREERKQL

```

```

DQFSRITSAYPDGRPTKRRDRQLDLRKKGDW"
complement(247675. .247860)
/genes="NMA1348"
/note="Pfam match to entry PF01479 S4, S4 domain, score
59.00, E-value 1e-13"
248099. .249058
/genes="accA"
248099. .249058
/genes="accA"
/BC_number="6.4.1.2"
/note="NMA1349, accA, probable acetyl-CoA carboxylase
carboxyl transferase subunit, len: 319 aa; similar to many
e.g. SW:ACCA_ECOLI (EMBL:M96394), accA, Escherichia coli
acetyl-CoA carboxylase carboxyl transferase subunit (EC
6.4.1.2) (318 aa), fasta scores: E(): 0, 64.9% identity in
313 aa overlap"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase carboxyl
transferase subunit"
/protein_id="CAB84596.1"
/db_xref="GI:7380018"
/db_xref="SPTREMBL:Q9JUF0"
/translation="MAPVFLDFEQIAELTNKIDELRFVQDESADVDSIEIHLQKKS
NDLTKSIFSKLTPAQISQVSRHPQRPYTLDTYIEALDFEELHGDHRAADDAIVGGGL
ARFNGSVNVGHQKGRDTEKIRRNFGMPRPGYRKALRLMKTAKEFGFLPVDYTFDIT
PGAYPGIGAERGQSAIGKNLYELTRLVPVLTIVIGGGGALAVAGDGVNNLQD
YSTYSVISPCCASILLWKAADAQAALGTADRLQKLDLVDVTYIKEPPLGGAHRDF
GQTMNVKAVLEKQLHEADQSIPLADLLSRFRDPRIMAYKFSEQ"
249156. .250436
/genes="NMA1350"
249156. .250436
/genes="NMA1350"
/note="NMA1350, len: 426 aa; unknown, similar to many
bacterial hypothetical proteins e.g. SW:MESJ_ECOLI
(EMBL:D49445), mesJ, Escherichia coli putative cell cycle
protein (432 aa), fasta scores: E(): 0, 34.3% identity in
428 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1350"
/protein_id="CAB84597.1"
/db_xref="GI:7380019"
/db_xref="SPTREMBL:Q9JUE9"
/translation="MKDCEPQGLNGKKTAVALSGGLDSVLLHLVRAGKKGGFIPDA
LHIHCLSPRADDWDFCONYCDMLGVLGYLETYKVCYKEKNGLCLEAAROKRYAAFAEK
GFDVLALAHRRDDQIETFMVARGGGLRALAMPVAFPEGEGIIWRLLPFSRODI
WDYAQKHGLPNIDESENTDTAYLRNFRHRIILPELSAQIPHFGHVLNNVRLAQEDLA
LLDEVVQDCRWVCGAGYEDTARWLTFSPRRKTHILRHLKENGIPVNPQNALADIAR
VLTEAKTGRNQLQGFELHHYAGRLVFRLKTDKRLFKDRQISGNRELLTQGGFVL
KRPFCGLPHELLLEODGILRTVAASDTLAMGGIHKDVKKILOGKRVLPLVRPIWPLVAD
SGNRPLALNCCADFOYSVSDGILPVHPDPFPLF"
complement(250388. .250397)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
250510. .251295
/genes="NMA1351"
250510. .251295
/genes="NMA1351"
/note="NMA1351, possible RNA methyltransferase, len: 261
aa; shows weak similarity to SW:TSNR_STRLU (EMBL:L39157),
tsnR, Streptomyces laurentii rRNA
(adenosine-2'-O)-methyltransferase (270 aa), fasta
scores: E(): 1.4e-07, 26.1% identity in 253 aa overlap.
Similar to many hypothetical proteins e.g. SW:YGY3_SYNY3
(EMBL:D90913), SLR1673, Synecocystis sp. (strain PCC
6803) hypothetical tRNA/rRNA methyltransferase (274 aa),
fasta scores: E(): 9e-20, 34.1% identity in 267 aa
overlap. Shows weak similarity to NMA1127, fasta scores:
E(): 1.3e-10, 27.7% identity in 224 aa overlap. Contains
Pfam match to entry PF00588 Spou_methylase, Spou rRNA
Methylase family"
/codon_start=1

```



```

/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(260300) .260304)
complement(260318) .262132)
/genes="cysJ"
complement(260318) .262132)
/genes="cysJ"
/EC_number="1.8.1.2"
/note="NMA1363, cysJ, probable sulphite reductase alpha
subunit, len: 604 aa; similar to many e.g. SW:CYSD_ECOLI
(EMBL:M23008), cysJ, Escherichia coli sulfite reductase
[NADPH] flavoprotein alpha-component (EC 1.8.1.2) (598
aa), fasta scores: E(): 0, 55.7% identity in 600 aa
overlap. Contains Pfam matches to entry PF00667
FAD-binding, FAD binding domain and to entry PF00175
oxidoreduced_fad, Oxidoreductase FAD/NAD-binding domain"
/codon_start=1
/transl_table=11
/product="putative sulphite reductase alpha subunit"
/protein_id="CAB84609.1"
/db_xref="GI:7380031"
/db_xref="SPTREMBL:Q9JUD8"
/translation="MSEHDMONTNPPPPPEITQLLSGLDAAQWALSGYAWAKAG
NGASAGLPAQTALPTAEPFSTVLSASOTGNASKSVADKAADSLEAGIQVSRAELKD
YKANIAGERRLLLVSTOGEPPPEAVVLLKLNKKAPKLDKLOFAVLIGLDSY
PNCIRAGKDFKFEELGAKRLLEFRVDALDFAAADGWTNDIAALLKEAANKRATP
APQTTTPAGLQAPDGRYCKADFPFPAALANQKITARSDKVRHIEIDLSGDLHYL
GDALGVFEDNPALVREILDLIDGIDQATEIQAGKTLFVASALLSHFELLRTQNTPAFV
KGYAPADDELDIRAADNAVILQGVQSTPIADVLRFPKLTAEQFAGLLRPLAPRL
YISSQAEVGVEMVLTGVAVRPEHEGRARGAGSGLADRLIEDCTVVRVVERNDGF
RLPDSRKPVTVMISGTVGAFRAFPVQRAAENAECKNMLFTGNPHFARDFLQYTEMQ
QPAKDFLHRYDPAWSRDOEEKIYQDKIREQAEGLQWQLQEGAHYVCGDAKMAKD
VENALLDIIGAGHLDDEGAEEYLDMEEREKRYQRQVY"
complement(260423) .260764)
/genes="cysJ"
/note="Pfam match to entry PF00175 oxidoreduced_fad,
Oxidoreductase FAD/NAD-binding domain, score 140.50,
E-value 2.9e-38"
complement(260846) .261433)
/genes="cysJ"
/note="Pfam match to entry PF00667 FAD_binding, FAD
binding domain, score 139.50, E-value 6.1e-38"
261441) .261450
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
261963) .261972
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(262129) .263415)
/genes="cysN"
complement(262129) .263415)
/genes="cysN"
/EC_number="2.7.7.4"
/note="NMA1364, cysN, probable sulphate adenylylate
transferase subunit 1, len: 428 aa; similar to many e.g.
SW:CYSD_ECOLI (EMBL:M74586), cysN, Escherichia coli
sulfate adenylylate transferase subunit 1 (EC 2.7.7.4) (475
aa), fasta scores: E(): 0, 46.8% identity in 432 aa
overlap. Contains Pfam match to entry PF00009 GTP_EFTU,
Elongation factor Tu family, PS00301 GTP-binding
elongation factors signature and PS00017 ATP/GTP-binding
site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative sulphate adenylylate transferase subunit
1"
/protein_id="CAB84610.1"
/db_xref="GI:7380032"
/db_xref="SPTREMBL:Q9JUD7"
/translation="MTAQHTPLRITAGSVDDGKSTLIGRLLYDSKALLSDQIKTLE
SGKSGKLETFIDSLDGLAEAREQGITIDVAYRYFSTAKRKFTIADTPGHEQVTRN
MYTGASTAAVILVDASQLDFGAQPLQLLPQTKRHSAILLRQLDCHPLVWVANKMDLL
DYSDEKFAIVEAYRRLAEQLGLKDAHFVPVSALLGDNIYVPGGNPKYKGEPLLSIL

```

```

ETLPGADEVSVRTADDFYPVOLVVRQADKADDFRGYQGRIGRSVAVGQTVRIEPLNG
LTAEVSEIITPKGEVAQAFAGEATRLDRDIDVSRGDLFVDKNPSLAPQKHLEATLIC
WFDERPLNTARKYVLLKHGTQTVPAKGEIEISVLVDRTLEQEGAESLKMNDIAKVRIN
LQKPVATPYAENTAAGSFILIDEATYGTVAAGMIL"
complement(262204) .263400)
/genes="cysN"
/note="Pfam match to entry PF00009 GTP_EFTU, Elongation
factor Tu family, score 279.00, E-value 6.3e-80"
complement(263185) .263232)
/genes="cysN"
/note="PS00301 GTP-binding elongation factors signature"
complement(263350) .263373)
/genes="cysN"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(263421) .263426)
complement(263454) .264377)
/genes="cysD"
complement(263454) .264377)
/genes="cysD"
/EC_number="2.7.7.4"
/note="NMA1365, cysD, probable sulphate adenylylate
transferase subunit 2, len: 307 aa; similar to many e.g.
SW:CYSD_ECOLI (EMBL:M74586), cysD, Escherichia coli
sulphate adenylylate transferase subunit 2 (EC 2.7.7.4) (302
aa), fasta scores: E(): 0, 48.2% identity in 301 aa
overlap. Contains Pfam match to entry PF01507 PAPS_reduct,
Phosphoadenosine phosphosulfate reductase family"
/codon_start=1
/transl_table=11
/product="putative sulphate adenylylate transferase subunit
2"
/protein_id="CAB84611.1"
/db_xref="GI:7380033"
/db_xref="SPTREMBL:Q9JUD6"
/translation="MTKTPENVLQDLWLESETHIIREVAACEENPALLFSGCKDSVV
LIAACKAFELGSRPVKLPPLVHIDTGHNYPEVIAFRDQAQAKLNALIVGRVDSI
AKGTVVRKRETSRNAQAQVLTLETTEANGFDALMGARDEDEKAKERIFSFDEF
GOWDPKAQPELWSLYNTRLKHGENMRVFPISNWTELDINQYIARNELEPPIYSHR
REYVRRGLLPVPTPLTPKMPSETSEILDVFRFTVGDICTCPVESTASTPTTEIIRT
AVADISERSATRLDDOASEAAMEKRKKEGVF"
complement(263655) .263664)
/genes="cysD"
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(263694) .264281)
/genes="cysD"
/note="Pfam match to entry PF01507 PAPS_reduct,
Phosphoadenosine phosphosulfate reductase family, score
-11.80, E-value 8.4e-06"
complement(264385) .264388)
complement(264415) .265149)
/genes="cysH"
complement(264415) .265149)
/genes="cysH"
/note="NMA1366, cysH, probable phosphoadenosine
phosphosulfate reductase, len: 244 aa; similar to many
e.g. TR:Q34620 (EMBL:AJ000974), cysH, Bacillus subtilis
phospho-adenylylsulphate sulfotransferase (233 aa), fasta
scores: E(): 2e-24, 37.1% identity in 205 aa overlap.
Contains Pfam match to entry PF01507 PAPS_reduct,
Phosphoadenosine phosphosulfate reductase family"
/codon_start=1
/transl_table=11
/product="putative phosphoadenosine phosphosulfate
reductase"
/protein_id="CAB84612.1"
/db_xref="GI:7380034"
/db_xref="SPTREMBL:Q9JUD5"
/translation="NKIFKPALWKIPPIENGSETALAEKTETLKQRLHRIAGSHRDPAR
FASLLAEDMVTIDLAGENLKITFTLDTGLTHAETLNLDRIERVYPHMOIKRFPQ
IREDALHVESKGRFAFYDVEARRECCRIKTEPLDRAIAGADWLTCORREOSATR
TELPAEYDAGRGGIDKYNPIEDWSEHWYAILANNVPYNDLYROGFPISGDCDCTPR
VKAGEDIRAGRWGIDKNKSEKGLHK"

```

misc_feature complement(264421. .265020)
/gene="cysH"
/note="Pfam match to entry PF01507 PAPS_reduct, Phosphoadenosine phosphosulfate reductase family, score 246.40, E-value 3.9e-70"
246.40, E-value 3.9e-70"
gene complement(265165. .266625)
/gene="cysG"
complement(265165. .266625)
CDS /gene="cysG"
/EC_number="2.1.1.107"
/note="NMA1367, cysG, sirohaem synthase, len: 486 aa; highly similar to TR:P95370 (EMBL:X10177), cysG, Neisseria meningitidis strain MC58 sirohaem synthase (EC 2.1.1.107) (475 aa), fasta scores; E(): 0, 97.5% identity in 474 aa overlap. Contains Pfam match to entry PF00590 TP_methylase, Tetrapyrrole (Corrin/Porphyrin) Methylases, PS00839 Uroporphyrin-III C-methyltransferase signature 1 and PS00840 Uroporphyrin-III C-methyltransferase signature 2."
/codon_start=1
/transl_table=11
/product="sirohaem synthase"
/protein_id="CAB84613.1"
/db_xref="GI:7380035"
/translation="MNYPFIFANLAGRPVLVVGGGAAVAAARKISLLIKAGAEVRVAAKH LNAELSALAAENKILWLAEFRAEHRVTFLIIAASSDOALNRRVFHFLAESQKPVNV VDRDHCSFIPSPVIDRNPVOIATVSSGSAPVLARLLRLEALLPSPSGDMAEISGR WRDAYGKGLKSVTERRRWEKQFNGRFAALVNQNTLAERLAGOLEQSRONDQGS VSLVAGPGDAGLLTLKLEIQDQADVLYDALVSDGISLVRRDAERIFVGRKRGG RFPQEDNALMVLRAEGRRVRLKGGDPFVFGRGEELETARHQIIPFVSPGITA VGTATYAGTPLTHRDYAQSAAVFTGHRKADAPDIEMQTLIARSQRTLIYMGALKAA LI AERLQOHRGSPDTPAAVISQGTLPAGKTATGTLANLAEALAEANPAPALIVIGEVLGH EKLAWFENAKKESNPAEHAYFALDGLGTGQEQQA"
complement(265339. .265974)
misc_feature /gene="cysG"
/note="Pfam match to entry PF00590 TP_methylase, Tetrapyrrole (Corrin/Porphyrin) Methylases., score 286.40, E-value 3.7e-82"
E-value 3.7e-82"
misc_feature complement(265636. .265737)
/gene="cysG"
/note="PS00840 Uroporphyrin-III C-methyltransferase signature 2"
265861. .265870
misc_feature /note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(265918. .265962)
misc_feature /gene="cysG"
/note="PS00839 Uroporphyrin-III C-methyltransferase signature 1"
complement(266633. .266637)
RBS complement(266898. .266907)
misc_feature /note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(266922. .267578)
gene /gene="NMA1368"
complement(266922. .267578)
CDS /gene="NMA1368"
/note="NMA1368, possible membrane protein, len: 218 aa; similar to TR:Q46471 (EMBL:U25811) Chromatium vinosum hypothetical protein (231 aa), fasta scores; E(): 1.1e-23, 37.9% identity in 211 aa overlap. Contains hydrophobic, possible membrane-spanning regions. C-terminus contains a poly-Ala tract"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84614.1"
/db_xref="GI:7380036"
/db_xref="SPTREMBL:Q9JUD4"
/translation="MKNKTKVMDFPTRLPFHLLAASLPFWMTYSAKTGGMDLQWHTRVG LFLIFLLVRLCWIWGSDTARFSRWGSGIREYMKNGIPEHVQPGNPGLGALMVV ALLAAVSFGVGTGLFAADNTSTNGYLNHLVSEHTGSLMRKIHNLNFFKLLAVFSADV

RBS IAAVAARYVKKKNLVLPMTGPKYIEGKTSIRFAGKAAALAAALSVAALAAAAILLLS
267634. .267638
gene 267643. .268533
/genes="NMA1369"
267643. .268533
CDS /gene="NMA1369"
/note="NMA1369, len: 296 aa; unknown, similar to the C-terminal half of bacterial hypothetical proteins e.g. SW:YJEF_ECOLI (EMBL:U14003), yJef, Escherichia coli hypothetical protein (515 aa), fasta scores; E(): 7.8e-28, 37.8% identity in 278 aa overlap. Contains Pfam match to entry PF01256 UPF0031, Uncharacterized protein family UPF0031"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1369"
/protein_id="CAB84615.1"
/db_xref="GI:7380037"
/db_xref="SPTREMBL:Q9JUD3"
/translation="MPPVPHLGSERRRMLQAAALRPHVFKAEDSHKGTFGTFLAV GGAVGSGAPVLAASAAAMVLCCKVWAGNQETLPFAVIAGFPEIMLDTADGLTKRQG INAWTAGCGLGTDKAAAVGTLGILTEHTDKPVLLDADALNLTSTDETRNLARGCKNL ILTPHPAAARLGLTTVAQVQADRTAAVRKIGAFGATVVLKGHKTULVAASDTEIAYVN ESGNAGLATAGSGDVLGGIIGSLLAQGVVFEACAGAWLHGAADVIKESAGIAAGL LAGEIAPAAARLNRNITESM"
267763. .268494
misc_feature /gene="NMA1369"
/note="Pfam match to entry PF01256 UPF0031, Uncharacterized protein family UPF0031, score 169.90, E-value 4.1e-47"
E-value 4.1e-47"
repeat_unit /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 270.84"
/label=Correia
complement(268713. .268722)
misc_feature /note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
268803. .270614
gene /gene="typA"
268803. .270614
CDS /gene="typA"
/note="NMA1370, typA, probable GTP-binding protein, len: 603 aa; similar to many e.g. SW:TYPA_ECOLI (EMBL:L19201), typA, Escherichia coli GTP-binding protein implicated in regulation of virulence factors, fasta scores; E(): 0, 64.1% identity in 599 aa overlap. Contains Pfam match to entry PF00009 GTP_EFTU, Elongation factor Tu family, PS00301 GTP-binding elongation factors signature and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein"
/protein_id="CAB84616.1"
/db_xref="GI:7380038"
/db_xref="SPTREMBL:Q9JUD2"
/translation="MKQIRNIIAHVDHGTTLVDQLLRSGTFRANQOQVDERVMDS NDLEKERGITILAKNTALDEYGVHINKIDRPSARPSWIDQTFELFNLGATDQLDLP GPMPQTRFVTKKALALGKPIVYHINIDRPSARPSWIDQTFELFNLGATDQLDLP IVYASGLSGFKLEETDESNDMRPLFDITLKYTPAPSGSADETLQIQSDXDNYTIS RLGIGIRLNGRIKPGQVAVNMHDQIQAQRINQLLGFGLERYLPEAEAGDITVIT QKEDIGIGVITIDKDNPKGLMPSVDEPTLTMDPMVNTSPLAGTEKFVTSRQIRDL QKELLTNVALRVEDTADADVFRVSGRGELHLTILLENMRREGYELAVGKRVVYRDID GQKCEPYENLTVDPDDNQCAVMEELGRRRGELTNMESDNGRLEHYHLPARGLTGF QGEFMLTRGVGLMWSHVDFDAPKPKPMGRHNGVLVSQGEQAVATALNWLDEGRKM FVSPNDKIGEMGIIGHSDNDLVNPNLKGKKLNTINIRASGTDEAVRLTTPTIKLTLEGA VFETIDDELVEITPQSIRLKRKYLSELERRRHFKKLD"
268806. .269981
misc_feature /gene="typA"
/note="Pfam match to entry PF00009 GTP_EFTU, Elongation factor Tu family, score 354.60, E-value 1.1e-102"
268833. .268856
misc_feature /gene="typA"


```

/db_xref="GI:7380041"
/db_xref="SPTREMBL:Q9JQW2"
/translation="MNIIMDSRKFGILWEENSECNGFIYQKTIITIGENIYKPIQCY
GYFTLNAVFNLSLKSPEEKYVAGGNGNLDGQQLDIDKYNSILCNIESIDFTYMSG
GNGCEDICLVLEMGYSGEERLFPYSPDNGKNEIRYKKGTVESVIFQLNL"
RBS
misc_feature
complement(276575..276578)
complement(276612..276621)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
misc_feature
276642..276651
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(276660..279218)
/genes="glnD"
CDS
complement(276660..279218)
/genes="glnD"
/EC_number="2.7.7.59"
/notes="NMA1374, glnD, probable [protein-PII]
uridylyltransferase, len: 852 aa; similar to many e.g.
SW:GLND_ECOLI (EMBL:M96431), glnD, Escherichia coli
[protein-PII] uridylyltransferase (EC 2.7.7.59) (890 aa),
fasta scores: E(): 0, 33.1% identity in 865 aa overlap"
/codon_start=1
/transl_table=11
/product="putative [protein-PII] uridylyltransferase"
/protein_id="CAB84620.1"
/db_xref="GI:7380042"
/db_xref="SPTREMBL:Q9JUC9"
/translation="MPANLSSALETFKQORDAAEHLKANRVSVFREYTAIVETLL
AALWAEYFNSALCLMAVGFGELFYPCSDVLAVVSPADSGIOEQIARVQTLW
DKCLMPSPGVSDELCSVRNDITGDTAFLEARFLFGNRQTAKLAERKMAQRNVAA
FVEAKLIVEMHRHAKSQGSQGVLEPNIKSCPGGLDIHTLLWIAKAQGLATDPLLK
ORLTTRAAGMLSHGYRRLAHIRIHLHLNAKRAEDRLFLDQPOVAESMGYEGNLRR
QSEELMRVYRAIKTVKQLSGILTPMLRSRVSSAPMRVTLRIDDDYLVQNNQIAARHT
DIFFRPEHIFIVEIMQORNDITALEPQTLRAWGATRKINRSFTQNSENRHPAGF
FRNGSLTQTLRNLNYGLVGRYPKAWKIVGLLQHDLFHIYVDVDDHILTVVNRVRL
ALDMHSHELFPYASALMQSFEQDILYLAAFFHDIAGRGGDHAIQGIADARQPAADF
LTGEESDLWLWVNLHLSAVAQKEDIQDPVDLDFCRVQVTHRLSLYLITISDI
RTNPKLWNAWRLSLESFLFAAGRYLTGNGNPHTLFGRRROEAADLLTRAAPPEKO
OKKLNALGSAYFARHQREILWHAANLVHDPETPIVRSRILFKSDSFQVMVPMNGP
RLFARLCRIFSRGHGFDILAARAFITEHDYILDTFIVQISQHPEDYPIQSLAEAL
NSFIHGTVAETQSHRRISRSRYMPIAPSTITPTPEEDYPWMSVEITAVNRPFLLA
DNAEVEFAHNSLVRYAKISTLDERAEDSTVFSLDLKNPKIQSSLKQTLLEQLS"
278198..278207
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
misc_feature
278952..278961
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(279267..279587)
/genes="NMA1375"
CDS
complement(279267..279587)
/genes="NMA1375"
/notes="NMA1375, possible transcriptional regulator, len:
106 aa; similar to e.g. regulatory proteins of restriction
modification systems e.g. TR:Q52622 (EMBL:M77223), pviIIC,
Proteus vulgaris regulatory transcription factor C (84
aa), fasta scores: E(): 1.8e-05, 42.2% identity in 64 aa
overlap. Similar to TR:AD39136 (EMBL:AF123569), nmeSIC,
Neisseria meningitidis strain 800615 (serotype B:4:PI.4)
putative control element (70 aa), fasta scores: E():
4.3e-05, 46.0% identity in 50 aa overlap. Contains pfam
match to entry PF01381 HTH_3, Helix-turn-helix. Contains a
probable helix-turn-helix motif at aa 36-57 (Score 1818,
+5.38 SD)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB84621.1"
/db_xref="GI:7380043"
/db_xref="SPTREMBL:Q9JUR09"
/translation="MMGNKLTLPALPELQDEQDLRAVLAYNMFLRVNKGWSQEELARQC
GLDRTVSAVERKRWNIALSNTEKMAALGVAAYQLLPPQERLKLMTNSADTRQMPs

```

```

ESGI"
complement(279279..279288)
/genes="NMA1375"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(279348..279512)
/genes="NMA1375"
/notes="Pfam match to entry PF01381 HTH_3,
Helix-turn-helix, score 49.60, E-value 6.9e-11"
complement(279911..280384)
/genes="bfrB"
complement(279911..280384)
/genes="bfrB"
/notes="NMA1376, bfrB, probable bacterioferritin B, len:
157 aa; similar to many e.g. SW:BFRB_NEIGO (EMBL:U76634),
bfrB, Neisseria gonorrhoeae bacterioferritin B (157 aa),
fasta scores: E(): 0, 99.4% identity in 157 aa overlap and
SW:BFR_AZOVI (EMBL:M83692), bfr, Azotobacter vinelandii
bacterioferritin (156 aa), fasta scores: E(): 0, 62.2%
identity in 156 aa overlap. Similar to NMA1377, fasta
scores: E(): 3.1e-17, 39.9% identity in 153 aa overlap"
/codon_start=1
/transl_table=11
/product="putative bacterioferritin B"
/protein_id="CAB84622.1"
/db_xref="GI:7380044"
/db_xref="SWISS-PROT:P56999"
/translation="MKGDRLVIRELNKLNLGLLTVINOVELHARILKNMGFEELGEHF
AFQSTIVEMKRAADDLIERILFLEGLPNLQELGKLLIGESTEEIACDTKQEKHEALL
AAATIAAQDQYVRDLLEKQDNEEHIDWLETOQLKIGLPNYLQTAQED"
complement(279923..280384)
/genes="bfrB"
/notes="Pfam match to entry PF01334 Bacteriofer.
Bacterioferritin, score 345.90, E-value 4.3e-100"
complement(280390..280394)
complement(280412..280876)
/genes="bfrA"
complement(280412..280876)
/genes="bfrA"
/notes="NMA1377, bfrA, probable bacterioferritin A, len:
154 aa; similar to many e.g. SW:BFR_A_NEIGO (EMBL:U76633),
bfrA, Neisseria gonorrhoeae bacterioferritin A (154 aa),
fasta scores: E(): 0, 99.4% identity in 154 aa overlap and
SW:BFR_PSEPU (EMBL:U66717), bfr, Pseudomonas putida
bacterioferritin (154 aa), fasta scores: E(): 0, 61.0%
identity in 154 aa overlap. Similar to NMA1376, fasta
scores: E(): 2.2e-17, 39.9% identity in 153 aa overlap"
/codon_start=1
/transl_table=11
/product="putative bacterioferritin A"
/protein_id="CAB84623.1"
/db_xref="GI:7380045"
/db_xref="SWISS-PROT:P56998"
/translation="MQGNQAVDVMNELLSGELAAARDQYFIHSLRYSEWYTKLFERL
NHMEETETHAEDFIIRILMGCTPKMARALNIGTDVSLRADLQTEYEVDRDLKK
GKLCLEADQYVSRDLMAVOLKDTEDHAHWLEQQLRLIELIGEGNYIQSQL"
complement(280415..280876)
/genes="bfrA"
/notes="Pfam match to entry PF01334 Bacteriofer.
Bacterioferritin, score 360.40, E-value 1.9e-104"
complement(280882..280887)
complement(281249..281258)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(281340..282323)
/genes="lipA"
complement(281340..282323)
/genes="lipA"
/notes="NMA1378, lipA, probable lipolic acid synthetase,
len: 327 aa; similar to many e.g. SW:LIP_A_ECOLI
(EMBL:L07636), lipA, Escherichia coli lipolic acid
synthetase (321 aa), fasta scores: E(): 0, 63.3% identity
in 308 aa overlap"

```


RBS
gene

286014. .286017
286028. .286687
/gene="ung"
286028. .286687
/gene="ung"
/EC_number="3.2.2.-"

/note="NMA1384, ung, probable uracil-DNA glycosylase, len: 219 aa; similar to many e.g. SW:UNG_ECOLI (EMBL:J03725), ung, Escherichia coli uracil-DNA glycosylase (EC 3.2.2.-) (28 aa), fasta scores; E(): 0, 67.0% identity in 215 aa overlap. Contains Pfam match to entry PF00315 UNG, Uracil-DNA glycosylase"

/codon_start=1
/transl_table=11
/product="putative uracil-DNA glycosylase"
/protein_id="CAB84630.1"
/db_xref="GI:7380052"
/db_xref="SPTREMBL:Q9JUC4"

/translation="MDTWHDALGGEKQPYFOEILNAVROERLSGGIIPPAADVENA
FRLTAFDRYKAVILGQDPYHGAGQAHGAFSVRQGIRIPPSLLNIYKELETDIEGFSI
PAHGCLTAWAEQVLLNTVLTVRAGQASHALLGWERFTDVIROLATHRKHLVPL
WGXYAQQRKRLIDSQNYLLITLTAHPSPLSAYRGFFGCRHESQANSYLSRHGIDPINWK
I."

misc_feature

286113. .286678
/gene="ung"

/note="Pfam match to entry PF00315 UNG, Uracil-DNA glycosylase, score 407.60, E-value 1.1e-118"

gene

286774. .289268
/gene="NMA1385"

CDS

286774. .289268
/gene="NMA1385"

/note="NMA1385, pseudogene, probable type I restriction-modification system modification protein, len: 2495 bp; similar to many e.g. SW:T1ML_ECOLI (EMBL:X13145), hsdM, Escherichia coli type I restriction enzyme EcoRI24II (320 aa), fasta scores; E(): 5.4e-11, 47.8% identity in 90 aa overlap. Highest similarity is to TR:Q9ZJ91 (EMBL:AE001564), HSDM_3, Helicobacter pylori J99 type I restriction enzyme (modification subunit) (815 aa), fasta scores; E(): 0.00044, 76.9% identity in 26 aa overlap (frame 3), E(): 7e-28, 62.3% identity in 146 aa overlap (frame 1), E(): 3.6e-18, 60.5% identity in 86 aa overlap (frame 3), E(): 1.4e-22, 65.2% identity in 132 aa overlap (frame 1), E(): 0.0009, 39.7% identity in 68 aa overlap (frame 3) and E(): 2.8e-12, 30.2% identity in 315 aa overlap"

/codon_start=1
/pseudo
/transl_table=11

/product="pseudogene, probable type I restriction-modification system modification protein"
complement(288743. .288752)

misc_feature

/note="Core DNA uptake sequence: gccgctctgaa"

gene

289395. .289505
/gene="NMA1390"

CDS

289395. .289505
/gene="NMA1390"

/note="NMA1390, possible pseudogene, len: 111 bp; shows weak similarity to parts of Helicobacter pylori hypothetical proteins e.g. TR:Q25951 (EMBL:AE000640), HPI401, Helicobacter pylori conserved hypothetical protein (235 aa), fasta scores; E(): 0.0001, 50.0% identity in 42 aa overlap. GC frameplot also suggests a short ORF here"

/codon_start=1
/pseudo

/transl_table=11
/product="hypothetical protein NMA1390 (pseudogene)"
complement(289837. .290352)

gene

/gene="NMA1391"

CDS

complement(289837. .290352)
/gene="NMA1391"

/note="NMA1391, possible periplasmic protein, len: 171 aa;

unknown, contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84633.1"
/db_xref="GI:7380053"
/db_xref="SPTREMBL:Q9JUC3"
/translation="MKSKLPLILNFSLISSPLGANAAKIYTCTINGETVYTTKPSKS
CLSTDLPPTGNSYSSERYIPQTSPTSPNGGOAVKYKAPVKTGSKPAKSNTPPQQ
APNSNRSRSLTETELSNERKALVEAQKMLSQARLKGNNHGEINQALQSNVLDROQN
IQALQRELGRM"
complement(290821. .292688)
/gene="NMA1393"
complement(290821. .292688)
/gene="NMA1393"
/note="NMA1393, pseudogene, possible ABC transporter ATP-binding protein, len: 828 bp; similar to many hypothetical ABC transporter proteins e.g. SW:YHES_ECOLI (EMBL:U19997), yhes, Escherichia coli hypothetical ABC transporter ATP-binding protein (637 aa), fasta scores; E(): 0, 56.8% identity in 345 aa overlap (frame 4) and E(): 0, 48.4% identity in 273 aa overlap (frame 6). A [cg]4 tandem repeat is present close to the frameshift site. Contains two Pfam matches to entry PF00005 ABC_tran, ABC transporter, two PS00211 ABC transporters family signature and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/pseudo

/transl_table=11
/product="putative ABC transporter ATP-binding protein (pseudogene)"
291005. .291014

/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS

complement(291221. .291230)
/gene="NMA1393"

/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS

complement(291229. .291645)
/gene="NMA1393"

/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 36.50, E-value 3.3e-08"
complement(291403. .291447)
/gene="NMA1393"

/note="PS00211 ABC transporters family signature"
complement(291669. .291676)
/note="(cg)4"

complement(292023. .292610)
/gene="NMA1393"

/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 150.70, E-value 2.5e-41"
complement(292197. .292241)
/gene="NMA1393"

/note="PS00211 ABC transporters family signature"
292399. .292408
/note="Core DNA uptake sequence: gccgctctgaa"

/label=DUS
292541. .292631

/note="termination of replication?"
/label=terminator?

complement(292566. .292589)
/gene="NMA1393"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
292746. .292755
/note="Core DNA uptake sequence: gccgctctgaa"

/label=DUS
complement(292765. .293160)

/gene="NMA1394"
complement(292765. .293160)

/note="NMA1394, len: 131 aa; unknown"
/codon_start=1

/transl_table=11

/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
298398. .299393
/gene="NMA1400"
298398. .299393
/gene="NMA1400"
/note="NMA1400, len: 331 aa; unknown, shows weak similarity to parts bacterial hypothetical proteins e.g. TR:CA849354 (EMBL:AJ248284) Pyrococcus abyssi hypothetical protein (197 aa), fasta scores: E(): 1.5e-06, 33.5% identity in 179 aa overlap. Contains PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1400"
/protein_id="CA884640.1"
/db_xref="GI:7380059"
/db_xref="SPTREMBL:Q9JUB9"
/translation="MKAAYQNIATVMQGFHNKQVLSAMKKVKRKPAAVRSDADIAA LFHAHRMQAQQSRILNMLAVEIRPGFVLDNKRAPDIRAALLDVXGEADGKPFPLN LLLGPMGAHEWHKGVAVPOLGGSITHVPGVFSPLRGEYLDLLAHAPSTGFOTADIG TGSGLAAILAKQGIPISVIGTDTNPKAVACARANTARLGFQKQVEIRETDLPEGFAD LIVCPNPLPAKPTSAVESALYDPESAMLAFLRDAPKHLNPDGEIRLISDLAEHLH LRPSDFLDKFAQAGLRVADVLTTRPVHKKAADPNPLAFARNRETTFLYRLKKA"
complement(298855. .298864)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
299055. .299075
/gene="NMA1400"
/note="PS00092 N-6 Adenine-specific DNA methylases signature"
complement(299465. .301210)
/gene="recD"
complement(299465. .301210)
/gene="recD"
/note="NMA1401, recD, exodeoxyribonuclease V alpha subunit, len: 581 aa; highly similar to TR:O68924 (EMBL:AF058330), recD, Neisseria gonorrhoeae exodeoxyribonuclease V alpha subunit (581 aa), fasta scores: E(): 0, 95.2% identity in 580 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="exodeoxyribonuclease V alpha subunit"
/protein_id="CA884641.1"
/db_xref="GI:7380060"
/db_xref="SPTREMBL:Q9JUB8"
/translation="MERQTFEQAARAARITRLEHYASSGDEVLANCTERLFAQLN GHSFIRLSGGEADALSALAPVVGTSAAPLILEGRRLFLGRMWQLEYDLAAEIKRLAAA DTSAPDAAGARONLAKWFGTSGEGORDAAALALQSFVITGPGTGKTTTVAKLLA LICGDENLPITALAAPTQKAAAHMARALHRAINGFDALAEVRRHLLKLEGOTVHRL LKRPKMQAENPVYPLPFDVLVIDEASMLDPTALMLQLLKAVKTGARVILLGDNQLP SVGIGAVLSVLSQKTVLDGETHQRLAGFLPEHGFSVSNPPVLAQNTAHLFSHRFGD NSGICLARAASVDEGAWALDFRPFDELEHSECSFNARVERLYRAHKAYQWQVKDGN IEAAYAGSDIIVYLAARQDAEDFNEAYCRHVRKRNINIPHELAYFAGEPIMIRQNDYA LELFNAGDILIMEDVGROGSLAAYPADADGFKKAVSCLPEPEPAFAMTVHKSGSEY REVWLLPSAAPSDEGDDALSGLSKELLYTAITRAREKFVFGGETFQQAATVKTR QTALCSMLERVESQG"
complement(300755. .300778)
/gene="recD"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(301216. .301219)
301224. .301262
/note="Stem loop containing DNA uptake sequences: caaat gccgtctgaa gccgtggggc ttcagacgc gcgg"
301229. .301238
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(301249. .301258)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(301278. .301973)
/gene="NMA1402"

CDS
complement(301278. .301973)
/gene="NMA1402"
/note="NMA1402, probable ABC-transporter ATP-binding protein, len: 231 aa; similar to many e.g. SW:GLNQ_BACST (EMBL:M61017), glnQ, Bacillus stearothermophilus glutamine transport ATP-binding protein (242 aa), Blastp scores: Expect 3.6e-29 and SW:YCFV_ECOLI (EMBL:AE000212), ycfV, Escherichia coli hypothetical ABC transporter ATP-binding protein (233 aa), fasta scores: E(): 0, 53.8% identity in 223 aa overlap. Contains pfam match to entry PF00005 ABC_tran, ABC transporter, PS00211 ABC transporters family signature and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative ABC-transporter ATP-binding protein"
/protein_id="CA884642.1"
/db_xref="GI:7380061"
/db_xref="SWISS-PROT:P57030"
/translation="MSELILKCEGVGKRYRDGGLDVLHLGLDEIRAGESTGIIGSS GSKSTPLLHILGGLDMPSEGRVLLMCGEDLRTLNORRLGDLNRHLCGFVQFHLLPPEF SALENVMPLLIGKKSREAEAAAMLEKVGGLKHRS THRAGELSGEQRQAARAL VTQPCLLADEPTGNLDNRANRVLDMLLELKTGLGVVYVTHDELAGEFRFVWMV KDGSHPKQGAN"
complement(301311. .301871)
/gene="NMA1402"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 209.80, E-value 4e-59"
complement(301494. .301538)
/gene="NMA1402"
/note="PS00211 ABC transporters family signature"
complement(301785. .301794)
/gene="NMA1402"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(301827. .301850)
/gene="NMA1402"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(301966. .303213)
/gene="NMA1403"
complement(301966. .303213)
/gene="NMA1403"
/note="NMA1403, possible integral membrane protein, len: 415 aa; unknown, similar to many bacterial hypothetical proteins e.g. SW:YCFW_ECOLI (EMBL:AE000212), ycfW, Escherichia coli hypothetical protein (414 aa), fasta scores: E(): 0, 35.3% identity in 405 aa overlap. Contains hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CA884643.1"
/db_xref="GI:7380062"
/db_xref="SWISS-PROT:P57061"
/translation="MPSLEAWIGLRYLRAKRRKNGFSFITMWSIAGIALGVTALIVVL SVYNGKQTEIRGOLLNVAPHAIEIGYIDNTDMRNLRTENRKRGITLAAAPVYSNOAL LANAGFIRGVQIRGILPSEERKVVEYDKMPACKGFEDLIPGEFDITLGVGLAALGAE VGNKVTIIPEGNVTPAGVYPRLKQPTVGLYKTVGYEVDNSLAMPHIODARVLYRLD KEVAGURLKADPQNAPALTALPEAQKQRTVVMRWTFNSRYSFYFAVELEKRMFI LLLITIAVAAFNLVSSLVMAVTEKQADIAILRTLGLSPGVKMFPMVOGAFSGFGFTLA GVVCGVLLGNVYGRVVAFFENLLGLVHINSQVYFIDYLPFDVDMGDAVIATICISLGIS FVATLYPSRRASTQPAEALRVE"
complement(301983. .301987)
/gene="NMA1403"
complement(302851. .302860)
/gene="NMA1403"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
303285. .303410
/gene="NMA1404"
303285. .303410
/gene="NMA1404"
/note="NMA1404, len: 41 aa; unknown"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1404"
/protein_id="CAB84644.1"
/db_xref="GI:7380063"
/db_xref="SPTREMBL:Q9JUB7"
/transl_table=11
/translation="MYRKQAASERPSAGDGAERAEHLNHNASNCPILSFRKTVS"
303409..303412
/gene="NMA1405"
/gene="NMA1405"
303419..303697
/gene="NMA1405"
/gene="NMA1405"
303419..303697
/notes="NMA1405, possible periplasmic protein, len: 92 aa;
unknown, contains a possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84645.1"
/db_xref="GI:7380064"
/db_xref="SPTREMBL:Q9JUB4"
/transl_table=11
/translation="MNIKRTAFALCAIALSAAAYAKAIKIDANNTPYSEADAKLAA
TAVGMGVKEPISLNGSGSITVSGSSNTQCQVFKGVNGALQIOGLNCK"
303708..303740
/notes="Stem loop containing DNA uptake sequences: aaat
gccgtctgaa ggc ttcagacggc atttt"
303713..303722
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(303726..303735)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(303753..304370)
/gene="recR"
/complement(303753..304370)
/notes="recR"
/notes="NMA1406, recR, possible recombination protein, len:
205 aa; similar to many e.g. SW:RECR_ECOLI (EMBL:X15761),
recR, Escherichia coli recombination protein (201 aa),
fasta scores: E(): 6.9e-27, 42.4% identity in 191 aa
overlap"
/codon_start=1
/transl_table=11
/product="putative recombination protein"
/protein_id="CAB84646.1"
/db_xref="GI:7380065"
/db_xref="SPTREMBL:Q9JUB6"
/transl_table=11
/translation="MSHKQDAFQGLIDALKVLPNVGPKSAQRIAYHLLQHRKRAEK
LVDALOTALKQVDHCARNTFCGGIDICADETRDRLMYVHPADYSNNMEAACH
DGLYFVLGQINTALGMVDSIALDRLAQRLGGGEVEEIIATATAGNATAYVLSE
FFKNLPYKVSRLSQGIPLGGELEYVDAGTLAQAVYERRLIKEGGA"
complement(304437..305975)
/gene="NMA1407"
/complement(304437..305975)
/gene="NMA1407"
/EC_number="5.2.1.8"
/notes="NMA1407, possible peptidyl-prolyl cis-trans
isomerase, len: 512 aa; similar to e.g. SW:CPDP_ECOLI
(EMBL:AE000150), ppib, Escherichia coli peptidyl-prolyl
cis-trans isomerase D (EC 5.2.1.8) (623 aa), fasta scores:
E(): 8.4e-14, 21.7% identity in 608 aa overlap. Contains a
probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative peptidyl-prolyl cis-trans isomerase"
/protein_id="CAB84647.1"
/db_xref="GI:7380066"
/db_xref="SPTREMBL:Q9JUB5"
/transl_table=11
/translation="MFHSIEKYRTPAOVLLGLIALTFVGGCVTSVHPGADYIVQVGD
EKISDHSINNNAIONEQADGGPSRDVAFQSLQRAYLKQGAQLMGISVSSSEQIKOII
DDPNFHANGKFDALLNRYLSQRHMSDEQVEETRDQFALQNLVNLVQNGVLVGDQ
AQKIDRQKQTVSETVKNAPFEERVAIPANEAKPSPFEQEAAYENELKKKAVADEN
KAKEKLGDDAFNHPSSIAEAANKSLKVTQETWLSRQDAQMSGPENLINAVFSDV

```

```

LKKKHNSVLTINSETAHVVRAKEVREBKTLPPFAEKADAVRQAVIRTEAAKLAENKAK
DVLTQLNGKXADVVKWSEVSVLGAQAPLAKAALQAQQAQANTFDLLIRYFNGKIKQTKGAQSDVN
GDGQ"
complement(306029..306403)
/gene="NMA1408"
/complement(306029..306403)
/gene="NMA1408"
/notes="NMA1408, len: 124 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YFEB_ECOLI (EMBL:X57403),
yffB, Escherichia coli hypothetical protein (118 aa),
fasta scores: E(): 5.8e-15, 38.5% identity in 117 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1408"
/protein_id="CAB84648.1"
/db_xref="GI:7380067"
/db_xref="SPTREMBL:Q9JUB4"
/transl_table=11
/translation="MILHIGPNCDDTVKKAKNRLAGYGLFEFRDFPKQKQPSAEICS
WLEQVPLATLLNKRGTSWRKLDAAETQOKVLSSTAFAVKLMSEMPSLIKRPVLECGGKV
YASFSEETDYGIFNRQAPCRQG"
complement(306287..306296)
/gene="NMA1408"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(306474..306483)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
306553..308181
/gene="NMA1409"
306553..308181
/gene="NMA1409"
/notes="NMA1409, probable ABC-transporter ATP-binding
protein, len: 542 aa; similar to many e.g. SW:TLRC_STRFR
(EMBL:M57437), tlrc, Streptomyces fradiae tylosin
resistance ATP-binding protein (548 aa), blastp scores:
Expect 1.4e-45 and SW:YBIT_ECOLI (EMBL:AE000184), ybit,
Escherichia coli hypothetical ABC-transporter ATP-binding
protein (530 aa), fasta scores: E(): 0.65, 6% identity in
543 aa overlap. Contains two pfam matches to entry PF00005
ABC_tran, ABC transporter and two PF00017 ATP/GTP-binding
site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative ABC-transporter ATP-binding protein"
/protein_id="CAB84649.1"
/db_xref="GI:7380068"
/db_xref="SPTREMBL:Q9JUB3"
/transl_table=11
/translation="MISTNGITMQFGAKPLFENVSVKFGEGNRYGLIGANGSKSTFM
KILGGLEQTAGEVAIEINGVRIGKLRQDFAYEDMRVLDVVMGHTEMMAAMTERDAI
YANPEATEDDYMAAELEAKFAEYDGYTAARAELLSGVIGSEDLHNATMAEVAPEGF
KLRLVLAQALFSKPDVLLDPTNLDINTIRWLEGLVNOYDSTWIIISHDRHFLNEV
CTHMADLDYNTIYPGNYDDYMLASASRRALAKAKAKELQELQFVAFVSANK
SKARQATSLKQADKIKSEMEVEFKSTQNPYIRFEADEKAKLHQRAVEVEKLARFE
TQLFKNLAFILAQORLAIIGNGAGKSTKLKLAGAYNPEYSDGLLPGDEGSIKWAKE
ASVGYTPQDHENDFDMDLSEMMQWQGDQEQVIRGTGLRLLGSGNDVYKVKVL
SGGEGRMLYGKLKLLKPNVLVMDPTNMDMESTESLNMALEKYNGLILFVSHDRQF
VSSLATQIIELDGKGVEHYLGDYESYLEKKGA"
306631..307236
/gene="NMA1409"
/notes="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 127.70, E-value 2.1e-34"
306652..306675
/gene="NMA1409"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
307403..307412
/gene="NMA1409"
/label=DUS
/complement(307594..308118)
/gene="NMA1409"
/notes="Pfam match to entry PF00005 ABC_tran, ABC

```

```
misc_feature
307615. .307638
/label=DUS
/note="NMA1409"
/notes="PS00017 ATP/GTP-binding site motif A (p-loop)"
308194. .308228
/notes="stem loop containing DNA uptake sequences: acaat
gcgctctgaa gcgcg ttcacagcgc attgt"
308199. .308208
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(308214. .308223)
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
308244. .308249
308253. .309506
/genes="cca"
308253. .309506
/genes="cca"
/EC_number="2.7.7.25"
/notes="NMA1410, cca, probable tRNA nucleotidyltransferase,
len: 417 aa; similar to many e.g. SW:CCA_ECOLI
(EMBL:M12788), cca, Escherichia coli tRNA
nucleotidyltransferase (EC 2.7.7.25) (412 aa), fasta
scores: E(): 0, 54.2% identity in 404 aa overlap"
/codon_start=1
/transl_table=11
/product="putative tRNA nucleotidyltransferase"
/protein_id="CAB84650.1"
/db_xref="GI:7380069"
/db_xref="SPTREMBL:Q9JUB2"
/translation="MOTYLVGAVRDYLLGLPVKDRDWWVGADQTMLAQGFQPVGK
DPVFLHPETHEEYALARTKAGYGVGFSPHADKDVTLQEDLMRRDLTINAAQDA
DGKIIDPGGORDLAAGILRHVSPAFADDPVRLTARFAARYKFEIAEETIKLRQM
VENGADALVAREVWQEFAGKGLMEKNPKRMIEVLECCGALKVLLPEVNALEFVGPORD
YHPEDSGIHTLMTLQRAADMGLSLPERYAALLHDLGAKTSPDILPRHHGHDLAGVE
PVRYNQDRAPKHCALAEALVCRWHIIFHQVGQKLSQTLNLVKKTKDAFRPERFOT
ALNVCIADTQRLNREHTPYQRAHWHLAFQWQKLSQTLNLVKKTKDAFRPERFOT
DRARLAQIAPLQKAPRAAQDKTEKH"
complement(309247. .309256)
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
309286. .309295
/genes="cca"
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
309528. .309589
/notes="Correia element; hmms hit to HMM Correia (1 - 62),
score: 99.50"
/label=Correia
309591. .309632
/notes="Correia element; hmms hit to HMM Correia (114 -
156), score: 56.34"
/label=Correia
309729. .310037
/genes="NMA1411"
309729. .310037
/genes="NMA1411"
/notes="NMA1411, len: 102 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1411"
/protein_id="CAB84651.1"
/db_xref="GI:7380070"
/db_xref="SPTREMBL:Q9JUB1"
/translation="MTLKTDLLPKINNEDYQRLIKHSAEFGSGGEIRLLNEILEKFNF
DVVQAAQAQAVMQQVRFDPNAYHIDSODEDTTGICPHCNPPMPLRLDYLVWRETRG
"
310060. .310102
/notes="Stem loop containing DNA uptake sequences: tcaat
gcgctctgaa acgcgcgcgacgc ttcacagcgc atacc"
310065. .310074
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
```

```
misc_feature
complement(310088. .310097)
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
310119. .311150
/genes="ruvB"
310119. .311150
/genes="ruvB"
/notes="NMA1412, ruvB, probable Holliday junction DNA
helicase, len: 343 aa; similar to many e.g. SW:RUVB_ECOLI
(EMBL:X07091), ruvB, Escherichia coli Holliday junction
DNA helicase (336 aa), fasta scores: E(): 0, 71.8%
identity in 333 aa overlap. Contains Pfam match to entry
PF00004 AAA, ATPases associated with various cellular
activities (AAA) and PS00017 ATP/GTP-binding site motif A
(p-loop)"
/codon_start=1
/transl_table=11
/product="putative Holliday junction DNA helicase"
/protein_id="CAB84652.1"
/db_xref="GI:7380071"
/db_xref="SPTREMBL:Q9JUB0"
/translation="MLOTDFNLTAQFQRTIVAQAQTASAOEELLERALKPKTLDDYIGOH
KAKEQLAIFIQAKKRGALDHVLLFGPGLGKTTLAHIIARELGNLRQTSQPVLER
AGDLAALLTNLDPHDVLFIIDEIHRFSPVVEEILYPALEDYRLDIMGEGPAARSVKID
LPPTLVGATTRAGMLTNPLRDRFGVSVLEEFYENRDLATIVRSQAQLQLDMSEEGA
ETIAKRSRGTPRIANRLLRRVRDFADVKNNGTIDGGIADAALSMLDVDVQGLDVMEDRK
FLEAVLHKFGGPGVGLDNVAAGIESTDITDIEVPYLIQQGLQRTPRGRMATERAY
LHFGLPVEK"
310314. .310853
/genes="ruvB"
/notes="Pfam match to entry PF00004 AAA, ATPases associated
with various cellular activities (AAA), score 5.00,
E-value 1.1e-05"
310329. .310352
/genes="ruvB"
/notes="PS00017 ATP/GTP-binding site motif A (p-loop)"
311152. .311193
/notes="Stem loop containing DNA uptake sequences: gcaat
gcgctctgaa acagagcctaatt ttcacagcgc atttc"
311157. .311166
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(311173. .311901)
/genes="rpe"
complement(311173. .311901)
/genes="rpe"
/EC_number="5.1.3.1"
/notes="NMA1413, rpe, probable ribulose-phosphate
3-epimerase, len: 242 aa; similar to many e.g.
SW:RPE_ECOLI (EMBL:Z19601), rpe, Escherichia coli
ribulose-phosphate 3-epimerase (EC 5.1.3.1) (225 aa),
fasta scores: E(): 0, 66.2% identity in 225 aa overlap.
Contains Pfam match to entry PF00834 Ribul_P3_epim,
Ribulose-phosphate 3 epimerase family, PS01085
Ribulose-phosphate 3-epimerase family signature 1 and
PS01086 Ribulose-phosphate 3-epimerase family signature 2"
/codon_start=1
/transl_table=11
/product="putative ribulose-phosphate 3-epimerase"
/protein_id="CAB84653.1"
/db_xref="GI:7380072"
/db_xref="SPTREMBL:Q9JUA9"
/translation="MTTYRIAPSILSADFARLGEEVESVIAAGADLIHFDVMDNHVVP
NLITFGPMVCAALKPYASVPIDVHLWVPEYDDLIQSFAKAGASILTFHPASPASHIDSL
SLIRDMGCGAGLVLPATPVYLLENVDLWLLMSVNPFGGSGSYFTYTLKIRIQV
RAMLDRYEKSGRRIATIEVDGGIKTDNIAAAQAQADTFVAGSAIFGKPDYKAVIDAM
RAELEKSGRLNLPAMIEIMPSEN"
complement(311179. .311188)
/genes="rpe"
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(311275. .311889)
/genes="rpe"
```



```

(EMBL:AJ242841) Neisseria meningitidis strain Z2491
putative transposase for IS1106A3 (335 aa), fasta scores:
E(): 0, 97.8% identity in 224 aa overlap (frame 6) and
E(): 0, 96.3% identity in 109 aa overlap (frame 1). The
potential coding region frameshifts at codon 109 approx.
near a cgg(2) repeat"
/codon_start=1
/pseudo
/transl_table=11
/product="putative transposase (pseudogene)"
/complement(325235. .325253)
/notes="inverted repeat at end of IS1106A3"
/label=IR
325279. .325313
/notes="stem loop containing DNA uptake sequences: aaaat
gcgctctgaa aggcct ttcagacgcc atttt"
325284. .325293
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(325299. .325308)
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(325348. .325662)
/gene="NMA1432"
complement(325348. .325662)
/gene="NMA1432"
/notes="NMA1432, pseudogene, possible type I restriction
enzyme system modification protein, len: 315 bp; similar
to parts of putative type I restriction enzyme system
modification components e.g. TR:Q9ZM08 (EMBL:AE001475),
HSDM_1, Helicobacter pylori J99 type I restriction enzyme
(modification subunit) (543 aa), fasta scores; E():
1.1e-07, 38.7% identity in 106 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative type I restriction enzyme system
modification protein (pseudogene)"
325681. .325835
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 284.48"
/label=Correia
complement(325932. .325941)
/notes="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
complement(325948. .327258)
/gene="NMA1433"
complement(325948. .327258)
/gene="NMA1433"
/notes="NMA1433, len: 436 aa; unknown, similar to many
hypothetical proteins e.g. SW:YCAJ_HAEIN (EMBL:U32833),
H11590, Haemophilus influenzae hypothetical protein (446
aa), fasta scores; E(): 0, 54.0% identity in 437 aa
overlap. may have some role in DNA replication or repair
as the N-terminal half is similar to yeast replication
factors e.g. SW:RFC3_YEAST (EMBL:L18755), rfc3.
Saccharomyces cerevisiae replication factor C 40 kD
subunit (340 aa), fasta scores; E(): 1.4e-10, 30.5%
identity in 233 aa overlap and to bacterial helicase
components SW:RUVB_THEMA (EMBL:U38840), ruvB, Thermotoga
maritima (334 aa), fasta scores; E(): 1.1e-07, 28.3%
identity in 244 aa overlap. Contains Pfam match to entry
PF00004 AAA, ATPases associated with various cellular
activities (AAA) and PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1433"
/protein_id="CAB84670.1"
/db_xref="GI:7380086"
/db_xref="SPTREMBL:Q9JU96"
/translation="MTDLFAREPYAPIAERLPRHTLDDVVQGEHLICEGKPLRVAVEG
GKPHSMLLGPPGVGKTTLARILAQSFNAQFLPVSASFVSGVKDIRAIDKAETALQQG

```

```

misc_feature
complement(326596. .327114)
/gene="NMA1433"
/notes="Pfam match to entry PF00004 AAA, ATPases associated
with various cellular activities (AAA), score 19.20,
E-value 8.8e-07"
complement(327076. .327099)
/gene="NMA1433"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(327264. .327267)
327507. .327992
/gene="NMA1435"
327507. .327992
/gene="NMA1435"
/notes="NMA1435, len: 161 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1435"
/protein_id="CAB84671.1"
/db_xref="GI:7380087"
/db_xref="SPTREMBL:Q9JU95"
/translation="MLIQSALCFEAHEAHEVNTIQEMEWVSIFSPQKDDSLITLKDE
KIYKNIYIVPWKKGENFRKLELITTKNDONGNLTVNGIYGISQFADPOLGYTPELSV
KSNNDILFNKLQSDVDPKPHSGNIHLDAKIQFNLVSDGDGVQDVASGKKYKQTIV
P"
327992. .328771
/gene="NMA1436"
327992. .328771
/gene="NMA1436"
/notes="NMA1436, len: 259 aa; unknown, rich in Ser, Thr and
Asn, lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1436"
/protein_id="CAB84672.1"
/db_xref="GI:7380088"
/db_xref="SPTREMBL:Q9JU94"
/translation="MRVNTISPPDENDKGSIKSPHSGTGNLFNMNVVHKDTHHFGIH
NLANTDVHLTAEDSDNVVAIKGELADPMPTGNVSTSLRGIEYAFKSNVTLTGKNNK
IISTAEKGGLGLGVNGLOSNSPYSGHTTGIFINNQSDDLTPQSLDGLSTIKLAKVGN
NEVDLNVKNHASVKGIIASHAKATLEAGDNIVQVRNPDNTALIDALKRQYPNASK
ELYRVGVQSTGTGTVRISPNKGKIALTELKTFGIQVNGICVSL"
complement(328767. .328921)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 276.86"
/label=Correia
complement(329043. .329957)
/gene="NMA1437"
complement(329043. .329957)
/gene="NMA1437"
/notes="NMA1437, probable lysR-family transcriptional
regulator, len: 304 aa; shows weak similarity to many e.g.
SW:XAPR_ECOLI (EMBL:X63976), xapR, Escherichia coli
xanthosine operon regulatory protein (294 aa), fasta
scores; E(): 1.6e-11, 23.7% identity in 295 aa overlap and
TR:O30441 (EMBL:AF006000), brg1, Bordetella pertussis
transcriptional regulator homologue regulated by the
oxygen-responsive transcriptional regulator btr (291 aa),
fasta scores; E(): 5.4e-13, 28.7% identity in 296 aa
overlap. Contains probable helix-turn-helix motif at aa
17-38 (Score 1810, +5.35 SD). Contains Pfam match to entry
PF00126 HTH_1, Bacterial regulatory helix-turn-helix
protein, lysR family and PS00044 Bacterial regulatory
proteins, lysR family signature"
/codon_start=1
/transl_table=11
/product="putative lysR-family transcriptional regulator"

```

```

RATILFVDEVHRNKAOQDAFLPHVESGLLTFTGATTENPSFEVNPALLSRAQYVILQ
PLSDDDLKLIKIALVLAPEYQETIETDAQLLVNTVDGARLLNLLEQLLRADTRF
RLNLTAEFLADSLGAQIRRFDEKGSFYQIISALHKSGSHPNALYAFRCMLDGG
TDPYLRARRIVTAWEDIGLADRALOINDAAATFERLKGSPGCELAAQAVLYLAA
AKSNAGYKAYNMHRHFVKENASDEVPVHLRNPTKLMKELGYGREYRYAHDEPNAYAA
GESYMPDGLDEPFYQPPVRGLEIKIGEKLAWLKSLEDEVLKAK"
complement(326596. .327114)
/gene="NMA1433"
/notes="Pfam match to entry PF00004 AAA, ATPases associated
with various cellular activities (AAA), score 19.20,
E-value 8.8e-07"
complement(327076. .327099)
/gene="NMA1433"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(327264. .327267)
327507. .327992
/gene="NMA1435"
327507. .327992
/gene="NMA1435"
/notes="NMA1435, len: 161 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1435"
/protein_id="CAB84671.1"
/db_xref="GI:7380087"
/db_xref="SPTREMBL:Q9JU95"
/translation="MLIQSALCFEAHEAHEVNTIQEMEWVSIFSPQKDDSLITLKDE
KIYKNIYIVPWKKGENFRKLELITTKNDONGNLTVNGIYGISQFADPOLGYTPELSV
KSNNDILFNKLQSDVDPKPHSGNIHLDAKIQFNLVSDGDGVQDVASGKKYKQTIV
P"
327992. .328771
/gene="NMA1436"
327992. .328771
/gene="NMA1436"
/notes="NMA1436, len: 259 aa; unknown, rich in Ser, Thr and
Asn, lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1436"
/protein_id="CAB84672.1"
/db_xref="GI:7380088"
/db_xref="SPTREMBL:Q9JU94"
/translation="MRVNTISPPDENDKGSIKSPHSGTGNLFNMNVVHKDTHHFGIH
NLANTDVHLTAEDSDNVVAIKGELADPMPTGNVSTSLRGIEYAFKSNVTLTGKNNK
IISTAEKGGLGLGVNGLOSNSPYSGHTTGIFINNQSDDLTPQSLDGLSTIKLAKVGN
NEVDLNVKNHASVKGIIASHAKATLEAGDNIVQVRNPDNTALIDALKRQYPNASK
ELYRVGVQSTGTGTVRISPNKGKIALTELKTFGIQVNGICVSL"
complement(328767. .328921)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 276.86"
/label=Correia
complement(329043. .329957)
/gene="NMA1437"
complement(329043. .329957)
/gene="NMA1437"
/notes="NMA1437, probable lysR-family transcriptional
regulator, len: 304 aa; shows weak similarity to many e.g.
SW:XAPR_ECOLI (EMBL:X63976), xapR, Escherichia coli
xanthosine operon regulatory protein (294 aa), fasta
scores; E(): 1.6e-11, 23.7% identity in 295 aa overlap and
TR:O30441 (EMBL:AF006000), brg1, Bordetella pertussis
transcriptional regulator homologue regulated by the
oxygen-responsive transcriptional regulator btr (291 aa),
fasta scores; E(): 5.4e-13, 28.7% identity in 296 aa
overlap. Contains probable helix-turn-helix motif at aa
17-38 (Score 1810, +5.35 SD). Contains Pfam match to entry
PF00126 HTH_1, Bacterial regulatory helix-turn-helix
protein, lysR family and PS00044 Bacterial regulatory
proteins, lysR family signature"
/codon_start=1
/transl_table=11
/product="putative lysR-family transcriptional regulator"

```

```

/misc_feature
/protein_id="CAB84673.1"
/db_xref="GI:7380089"
/db_xref="SPTREMBL:Q9JRJ3"
/translation="MDAVQLKSFVAHAEHGNLTQAAKRLFLSQPAVSAQIKALEEYVG
TPLEFRRTGKGMVLRAGELILLPEAESLLOYKHKLEHFAKTLAGDYSEETSLGIITHPID
SAKLVALDNIQTGAPKTRLRHIOXGMSGEILSRIOHKTLLHGFTLGNAAORGIRSVFL
QNLTYALICPOSOXPHLTRSLPOSLOECVWLEMSGVSGSRKHLHQFWRNRLSPKKQI
LCDYPTQIIDLAVAGIGVAMVPGNKAAEAKAGAGVALLIESCRHSMPLNFIYAEYED
NPHVSLLECEIEKRVGQVAVOPPVSDN"
/complement(329523..329951)
/gene="NMA1437"
/notes="Pfam match to entry PF00126 HTH_1, Bacterial
regulatory helix-turn-helix protein, lysr family, score
117.10, E-value 3.4e-31"
/misc_feature
/complement(329817..329909)
/notes="NMA1437"
/notes="PS00044 Bacterial regulatory proteins, lysr family
signature"
RBS
/complement(329963..329968)
RBS
330201..330206
gene
330213..331682
/gene="NMA1438"
330213..331682
CDS
/gene="NMA1438"
/notes="NMA1438, possible integral membrane protein, len:
489 aa: unknown, contains hydrophobic, possible
membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84674.1"
/db_xref="GI:7380090"
/db_xref="SPTREMBL:Q9JU93"
/translation="MEKHNGTYRDLHRPASEFATRDYLEHLELQIMQPKWRPNLPER
DYRFEWEDLIPAMAGTIGKVMVGAVAAFAAPLGPDSFVLENRYVELLIAAAFIILL
VSGFFPLPGANLPGTHGPLIPMIPVVSAGGHPLAFGISIAVLGLMLALFRGGSTMAKL
WSGVCGLLLYLGFICTGQVKLFESWAGSNMPTAFVLIIVTYMYALLHSHKRR
TLAVPLGLLAGVAAPFALGAPFEHPTAPGLPMSPAYWNGENSGHGLPTAESFLVV
FPFVALVAMWSDFLGHQVFQKLSYPERKYDKVLMNIDDTMTSCVRAQVGSILGGAN
FTSWGTYIVPASIAKRPPIPGAVLTAVLCLIIAGLWGPMDLAIWQPLVSVALVGVY
LPLEAGMENTRKGTQSAATVFSALVNPVFGWALTMLLDNLGLIGKERSAQLG
FTGRPLIPVVGFLILCVAMGAVMLPGIPPLEHFKSLG"
/complement(331713..331722)
/notes="Core DNA uptake sequence: gccgtctgaa"
/misc_feature
/label=DUS
BASE COUNT 79079 a 81425 c 89192 g 82105 t
ORIGIN
NMA422491 Length: 331801 February 15, 2002 13:45 Type: N Check: 8990 ..
Found using 'seq4' (pappu403.key)
...
39507 AACAAATACGGTTGAGCTGTGCGGAGAACCGCGACAGTGGGGGCAAGTCGGAACGTGG
|-----|
39557 39564
39567 AACTGAATAACGGAATTTGAACATCAGCTGGTGTTCACCGAGATTG
|-----|
...
82968 ATATACTGCACGAGATTTTGGCGGCAATCAGGCGCCTTGACCAGTTGTCGCGAACGTTGC
|-----|
83018 83025
83020 83027
83028 GCGCGGAACCCCAATCGGTGCTGAGCTGGCGACAAAGAATTTTGCCTG
|-----|
...
90541 AAAAGGCGTATTGTGTCGCAATATATGCTGGCAGCGGTAGCGCGGTCCGACGTTGA
|-----|

```

```

90601 CAATATCGTGTGGAAAAACATCAATATCGAAAAAGATCGCGCGCGGT
90591 90598
...
151536 ATTCTTCTGAAGAAACCGGAACTGATATTCACATCGTCAAAACGCTGCGCGAACGTTTG
|-----|
151586 151593
151596 CCCGCACATATTCCGCTCACCGCAAAAATGCGGCTCGGCTATGAAGAC
...
152016 AATAATACCGCATCGCCCGTCTTAAGCAATGTTTGGGCATCATGAAAAAACGTTCCAC
|-----|
152066 152073
152076 CCGGGCGCAACGCTGTTTGACCGCGTCCGAACGGTTAAGGATGCGGAT
...
173540 GGCTTTGAGTAGTGGTGTGCAAGGAGCCCGATAAAGAATAAGCGCGCAACGTTCCGCC
|-----|
173590 173597
173600 ACCGACAACGGATTTCGAGCAGGTAGGTATAGGGCGGTTTGGCGAGTTT
...
195237 AGTACGACCGCGCTGCCGCTTTTTCAGTGCCAAAGCCGCGCATCGGACGTAAACGTTCCGG
|-----|
195287 195294
195297 CGGCTTCGTAAATAATGTCGCGATAACGCCCATCGCCAGCGCTTTTGT
...
199172 CTCGACTGGCTGCCCGAAGTGTGTGGCGACGAATTCGCAATTTACCGCGCGCGCAACGTTCC
|-----|
199222 199229
199232 GAATTCGAACGGCTGTTTGGGCAATGCGGATGGAACAACCCGCGCG
...
244356 TGTAAAGCAATTTTGAAGGCAGCAATGCGCCATTTTGGAAAAACACGCTATCGCAACGTTCC
|-----|
244406 244413
244416 CGGATTGGAACGCATCCTCGCGTCAAAACCGACAATTACAAAGAAAA
...
-----
1 match found in sequence:
ttr237770 ; TOIG of: ttr237770 check: 3819 from: 1 to: 345
(from "immun_ge.seq")
TOIG of: ttr237770 check: 3819 from: 1 to: 345
LOCUS TTR237770 345 bp mRNA INV 01-JUL-1999
DEFINITION Trichuris trichiura mRNA for macrophage migration inhibitory
factor-like protein.
ACCESSION AJ237770
VERSION AJ237770.1 GI:5327285
KEYWORDS macrophage migration inhibitory factor-like protein; MIF gene.
SOURCE Trichuris trichiura
ORGANISM Trichuris trichiura
REFERENCE Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida;
1 (bases 1 to 345)
AUTHORS Tan,T.H.P. and Meyer,D.J.
TITLE Molecular cloning and expression of cDNAs encoding for novel

```

macrophage migration inhibitory-like proteins from the parasitic nematodes *Trichinella spiralis* and *Trichuris trichiura*
Unpublished
REFERENCE 2 (bases 1 to 345)
AUTHORS Tan, T.H.P.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1999) Tan T.H.P., Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UNITED KINGDOM
REMARK Revised by author 19-MAY-1999
FEATURES Location/Qualifiers
source 1..345
/organism="Trichuris trichiura"
/db_xref="taxon:36087"
/country="Jamaica"
/dev_stage="adult stage"
1..345
/gene="mif"
1..345
/gene="mif"
/function="potential immunomodulator"
/codon_start=1
/transl_except=(pos:19..21,aa:Ser)
/product="macrophage migration inhibitory factor-like protein"
/protein_id="CAB46355.1"
/db_xref="GI:5327286"
/translation="MPIFTSTNPSENLSVDLKLSTKLIAGMLGKPSYVAVHNG GKRIFFGGTDAPAGFQQLSLGSGVGEKRNRSKLFKHLTDGLGIPGNRMVINFVDM RGSVDVINGSTP"

BASE COUNT 85 a 82 c 85 g 85 t 8 others

ORIGIN
TTR237770 Length: 345 February 15, 2002 13:46 Type: N Check: 3819
Found using 'seq4' (pappu403.key)

1 ATGCCWATYTYACRTTYNSNAGCAAGTTCCTTCTGAGAACATTTCCGTCGATTCTCTG
23 30
61 AAGAGCACACAGCAAGTTGAT
...
2 matches found in sequence:
yppcp1 : TOIG of: yppcp1 check: 5414 from: 1 to: 9612
(from "immun_ge.seq")
TOIG of: yppcp1 check: 5414 from: 1 to: 9612

LOCUS yppcp1 9612 bp DNA BCT 19-AUG-1999
DEFINITION Versinia pestis plasmid pPCP1.
ACCESSION AL109969
VERSION AL109969.1 GI:5763810
KEYWORDS coagulase; fibrinolysin; IS100; omptin; pesticin; pesticin immunity; pim; pla; pst; rop; transcriptional regulator.
SOURCE Versinia pestis.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Versinia.
1 (bases 1 to 9612)
AUTHORS Karlyshev,A.V. and Wren,B.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 9612)
AUTHORS Baker,S.G. and Mungall,K.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 9612)
AUTHORS Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1999) Versinia pestis sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA
E-mail: barrell@sanger.ac.uk DNA supplied by Dr. Andrey Karlyshev and Prof. Brendan Wren, [3]. Department of Infectious and Tropical

COMMENT

Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT
Notes:
Versinia pestis sequencing at The Sanger Centre is funded by Beowulf Genomics
Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/Y_pestis/)
CDS are numbered using the following system eg yppcp1.01c. yp (Y. pestis), PCP1 (plasmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon..

FEATURES Location/Qualifiers

1..9612
/organism="Yersinia pestis"
/plasmid="pPCP1"
/strain="CO-92 Biovar Orientalis"
/db_xref="taxon:632"
1..1954
/note="IS100"
/label="IS100"
87..1109
/gene="YppCP1.01"
87..1109
/gene="YppCP1.01"
/note="YppCP1.01", probable IS100 transposase, len: 340 aa; identical to TR:p74993, and similar to e.g. TRA6_BACST putative transposase for insertion sequence element IS5376 (400 aa), fasta scores; opt: 632 z-score: 1097.4 E(): 0, 38.1% identity in 328 aa overlap. Contains helix-turn-helix motif from aa 19-40 (Score 2045, +6.15 SD)
/codon_start=1
/transl_table=11
/label=YppCP1.01
/product="putative transposase"
/protein_id="CAB53164.1"
/db_xref="GI:5763811"
/translation="MVTFTVMEIKILHKQGMSSRAIARELGISRNTVKRYLOAKSEP PKYTPRAVASLLDVRDYRDIORADIADHPYKIPATVIAEIRDQYRGGMTILRAFIR SL5VPCQEPAPVRFETEPGRQOVDMGTNRGSRPLHVFVAVLGYSRLYIEFTDNMR YDTLETCHRNAFRFGVGVPREVLYDNMKTVVLORDATQGHQHPFSLWQFGKMGFS PRLCFRACQTGKVMVQYTRNSYIPLTRPMGTIVDVETANRHGLRWLHBDVA NORKHETIQARPCDRWLEEQSQSLALPPEKKEYVDHLDENLVNFDKHLPHLPLSIYDS FCRGVA"
1106..1888
/gene="YppCP1.02"
1106..1888
/note="YppCP1.02", IS100 ATP-binding protein, len: 260 aa; identical to TR:p74994 and similar to e.g. IS7B_ECOLI insertion sequence IS21 putative ATP-binding protein (265 aa), fasta scores; opt: 730 z-score: 1078.9 E(): 0, 47.4% identity in 249 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop)
/codon_start=1
/transl_table=11
/label=YppCP1.02
/product="putative ATP-binding protein"
/protein_id="CAB53165.1"
/db_xref="GI:5763812"
/translation="MMMLQHQRLMALGQLQLESLSAAPALSQAQVDPQEWSMDFL

108 aa overlap, and TR:053332 (EMBL:AL021646) MT014.26
(RV3182) (114 aa), fasta scores; opt: 214 z-score: 320.5
E(): 1.5e-10, 39.4% identity in 104 aa overlap"
/codon_start=1
/transl_table=11
/label=YPPCP1.09c
/product="hypothetical protein"
/protein_id="CAB53172.1"
/db_xref="GI:5763819"
/translation="MMVLFSORFDDMLNEOEDALQEKVLADLKLVQVGPPLPRPYA
DTVKGSRYNMKELRVQFSGRPIRAFYAFDPIRRRAIVLCAGDRSNDRRFYEKLVRIAE
DEFTAHLTLESK"
misc_feature
8528..8532
/notes=diff: AF053945 has 3 T's this sequence has 5"
/label=diff

BASE COUNT 2792 a 2253 c 2098 g 2469 t
ORIGIN
YPPCP1 Length: 9612 February 15, 2002 13:46 Type: N Check: 5414 ..
Found using 'seq4' (pappu403.key)

1274 CAACGTAAACAGGCGATGTATACCCGAATGGCAGCCTTCCCGCGGTGAACGTTCCGAA
1324 1331
1334 GAGTATGACTTCACATTCCGCCACCGGAGCACCAGCAAGCAACTCCAG
1694 AAGAGCGCAATGATCCTGACATCCAATCTGCCGTTCCGGCAGTGGGATCAACGTTCCGCC
1744 1751
1754 GGTGATGCAGCACTGACCTCAGCGATGCTGGACCGGTATCTTACACCAC

-- Search Statistics --
Times: CPU 00:01:21.16 Total Elapsed 00:01:23.00
Number of sequences searched: 2013
Number of sequence hits: 24
Number of separate matches: 70
Number of sequence hits saved: 0


```

/misc_feature
/feature="NMA1350"
/feature="NMA1350", len: 426 aa; unknown, similar to many
bacterial hypothetical proteins e.g. SW:MESJ_ECOLI
(EMBL:D49445), mesj, Escherichia coli putative cell cycle
protein (432 aa), fasta scores; E(): 0, 34.3% identity in
428 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1350"
/protein_id="CAB84597.1"
/db_xref="GI:7380019"
/db_xref="SPTREMBL:Q9JUE9"
/translation="MKDCFPOGLNKKTAVALSGGLDSVLLHLLVRAGKKGGRIPDA
LHIIHGLSPRADWDQCNQYCDMLGVGLETVKVCYKNGLGIEAAAKQKRIAAFAEK
GFDVALAHRDDQIETFLAVARGGRLRAAPVAPFGEKGIWRLLPFRQDI
WDYAKGLPLNIEEDNTDYLNRFRHRLPELSAQIPHGRHVLNVRALQEDLA
LLDEVVQDCRWCCAGYFDTARWLTFSPRKRTHLRHFLKENGIPVPMQALADIAI
VLTEAKTGRNMLQGLFELHHVAGRLTFVLEKTDKRLFKDROIISGNLREIITGGFVL
KRHPFGLPEHLLEODGILRTVAASDTLAMGGIHKDKKILQGRKRLVPLRIPWLVAD
SGNRPALANCCADFOYSVSDGILPVHPDFILF"
complement(250388..250397)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
250510..251295
/feature="NMA1351"
250510..251295
/feature="NMA1351"
/feature="NMA1351", possible RNA methyltransferase, len: 261
aa; shows weak similarity to SW:TSNR_STRLU (EMBL:L39157),
tsnr, Streptomyces laurentii rRNA
(adenosine-2'-O-)-methyltransferase (270 aa), fasta
scores; E(): 1.4e-07, 26.1% identity in 253 aa overlap.
Similar to many hypothetical proteins e.g. SW:YG73_SYNY3
(EMBL:D90913), SLR1673, Synecocystis sp. (strain PCC
16803) hypothetical tRNA/rRNA methyltransferase (274 aa),
fasta scores; E(): 9e-20, 34.1% identity in 267 aa
overlap. Shows weak similarity to NMA1127, fasta scores;
E(): 1.3e-10, 27.7% identity in 224 aa overlap. Contains
pfam match to entry PF00588 Spou_methylase, Spou rRNA
Methylase family"
/codon_start=1
/transl_table=11
/product="putative RNA methyltransferase"
/protein_id="CAB84598.1"
/db_xref="GI:7380020"
/db_xref="SPTREMBL:Q9JUE8"
/translation="MKHISNTNNEHRLHRLLSQGRFRQYAOVTVLEGVHLLQVFLQ
SGMPVGVYIPEAKMPSSEVRKLTAVLPEDGFFSVSDGILKISLTCAADDVALIDI
PDAGALPAGGCVVLDGVQDPGNVTVLSRAAAGIGAVILGKCADAWSPKVLIRAGM
GAHFLSEIYPOADLEIWLRYKGRVFATLREEKQAVLYGEDLCEPTANWPGNEGAGV
GKAVLDRADKCVRIPMHDATESLNVAAMAATLCLFEQMRQRAAY"
complement(250579..250588)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
250686..250695
/feature="NMA1351"
/feature="NMA1351", possible RNA uptake sequence: gccgtctgaa"
/label=DUS
/feature="Core DNA uptake sequence: gccgtctgaa"
complement(250746..250755)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
250843..251265
/feature="NMA1351"
/feature="NMA1351", possible RNA uptake sequence: gccgtctgaa"
/label=DUS
/feature="Core DNA uptake sequence: gccgtctgaa"
complement(251307..251316)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
251346..251582
/feature="NMA1353"
251346..251582
/feature="NMA1353", possible lipoprotein, len: 78 aa; unknown,

```

```

contains a possible N-terminal signal sequence and an
appropriately positioned PS00013 Prokaryotic membrane
lipoprotein lipid attachment site, therefore may be a
lipoprotein"
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAB84599.1"
/db_xref="GI:7380021"
/db_xref="SPTREMBL:Q9JRJ5"
/translation="MYRYAAVAVFLCGCFEAQEGGIAEGRVETHLIVQCRGGGFGF
LGIVVWVLTDTAQNPAANGIRTNKEEGKFKVSTN"
251358..251390
/feature="NMA1353"
/feature="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
251704..252117
/feature="NMA1354"
251704..252117
/feature="NMA1354"
/feature="NMA1354", len: 137 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1354"
/protein_id="CAB84600.1"
/db_xref="GI:7380022"
/db_xref="SPTREMBL:Q9JUE7"
/translation="MGSLONRGYKIANRLKCGDAMKNCFFYDPTRIYDSGADYLTR
EKHLRVIANSAWGLLLNLSCYDEVLKRIPIGKQEIDDMDKVSALKRKFNDISE
IKVGDGWEYFPNPGQMKELDEVLLKYIPFEER"
252118..252123
252129..252434
/feature="NMA1355"
252129..252434
/feature="NMA1355"
/feature="NMA1355", possible periplasmic protein, len: 101 aa;
unknown, contains a possible N-terminal signal sequence.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84601.1"
/db_xref="GI:7380023"
/db_xref="SPTREMBL:Q9JUE6"
/translation="MRVSKIIGSMLLVTAVQTVFSANVYACRHNGKTSYSQTPGKHCT
NAGLGRDLVYSVRPAVKDRAEDAGVDYSDTVRDEHVONPRENTHKDSANTGAKTH"
252204..252227
/feature="NMA1355"
/feature="PS00017 ATP/GTP-binding site motif A (P-loop)"
252496..253872
/feature="mpl"
252496..253872
/feature="mpl"
/feature="mpl"
/feature="mpl"
/feature="NMA1356, mpl, possible
UDP-N-acetylmuramate:L-alanyl-
gamma-D-glutamyl-meso-diaminopimelate ligase, len: 458 aa;
similar to SW:MPL_ECOLI (EMBL:U14003), mpl, Escherichia
coli UDP-N-acetylmuramate:L-alanyl-
gamma-D-glutamyl-meso-diaminopimelate ligase (EC 6.3.2.-)
(457 aa), fasta scores; E(): 0, 57.3% identity in 457 aa
overlap. Shows weak similarity to NMA2061, fasta scores;
E(): 1.5e-19, 28.4% identity in 479 aa overlap. Contains
pfam match to entry PF01225 Mur_ligase, Mur ligase family"
/codon_start=1
/transl_table=11
/product="putative
UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-
diaminopimelate ligase"
/protein_id="CAB84602.1"
/db_xref="GI:7380024"
/db_xref="SPTREMBL:Q9JUE5"

```

misc_feature
/gene="NMA1343"
/note="NMA1343, len: 186 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1343"
/protein_id="CAB84591.1"
/db_xref="GI:7380013"
/db_xref="SPTREMBL:Q9JUF2"
/translation="MWIDDIYFDGLYIYGVLNNEPGLTNVEOGESVCVPVDDISDNM
FVNGIPYGGFTIQAMRGOMTEERTEHDAWGIDGDPGVLLVYEKEHPENLEEH
PAGNCIDDRQOLSQNPDLFHEQDEGTYPLHHEATAGNALMWQAMLEYGANPASKT
SEGYTALDFARLAGQNVADLLEPRH"
/codon_start=1
/gene="NMA1343"
/note="Pfam match to entry PF00023 ank, Ank repeat, score
27.90, E-value 0.00024"
245793..246134
/gene="fdx"
245793..246134
/gene="fdx"
/note="NMA1344, fdx, probable ferredoxin, len: 113 aa;
similar to many e.g. SW:FER_ECOLI (EMBL:M88654), fdx,
Escherichia coli ferredoxin 2Fe-2S (110 aa), fasta scores:
E(): 1.2e-30, 67.3% identity in 110 aa overlap. Contains
Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur
cluster binding domains"
/codon_start=1
/transl_table=11
/product="putative ferredoxin"
/protein_id="CAB84592.1"
/db_xref="GI:7380014"
/db_xref="SPTREMBL:Q9JOT7"
/translation="MPKITVLPHTLCPEGAVIDNAPEGKTVLDVLLDHPDIEVDHACE
KSCACTCTCHVIRKGFDSLEETLEEDLDQAWGLEADSRSLSCQAVVAGEDLIVEIP
KTYINHAREEH"
245799..246089
/gene="fdx"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 104.20, E-value
2.5e-27"
246136..246175
/note="Stem loop containing DNA uptake sequences: aacag
gccgtctgaa gccgcgcgc ttcagacgc attgt"
246141..246170
246141..246150
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(246161..246170)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
246236..246296
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 81.29"
/label=Correia
246298..246340
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 44.42"
/label=Correia
complement(246421..247008)
/gene="NMA1345"
complement(246421..247008)
/gene="NMA1345"
/note="NMA1345, len: 195 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1345"
/protein_id="CAB84593.1"
/db_xref="GI:7380015"
/db_xref="SPTREMBL:Q9JUF1"
/translation="MNNMFAAKLSKLVYVTASDHPDLSLSEMEFPDRILLIRKIYQILD
GQHLSRVTVCLHHNRRGLIALDKAAGCDAAMLRAQRWRYTQIAERLSGGSGFT
VTAESVSAACPELEGRLLEIVRAAVSFGTQGOENPAGGQENTQAVPDAASGHSVRK
PIKINIEAKVERYPAYRPTAALKVTRRRFRLLQNR"

repeat_unit
complement(247004..247139)
/note="REP 2; hmms hit to HMM REP 2 (1 - 128), score:
129.26"
complement(247013..247018)
247221..247418
/gene="NMA1347"
247221..247418
/gene="NMA1347"
/note="NMA1347, len: 65 aa; unknown, similar to bacterial
hypothetical proteins e.g. TR:069223 (EMBL:AF010139)
Azotobacter vinelandii hypothetical protein (68 aa), fasta
scores: E(): 2.9e-14, 56.9% identity in 65 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1347"
/protein_id="CAB84594.1"
/db_xref="GI:7380016"
/db_xref="SPTREMBL:Q9JR02"
/translation="MKWTDQRTAEELYDLHGETIDPRTVFTQLRLIMALPEFDD
PARCGERILEAVQQAIDEAE"
247430..247439
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(247486..247887)
/gene="NMA1348"
complement(247486..247887)
/gene="NMA1348"
/note="NMA1348, len: 133 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YRFH_ECOLI (EMBL:U18997),
yrfH, Escherichia coli hypothetical protein (133 aa),
fasta scores: E(): 1.3e-19, 45.5% identity in 132 aa
overlap. Contains Pfam match to entry PF01479 S4, S4
domain"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1348"
/protein_id="CAB84595.1"
/db_xref="GI:7380017"
/db_xref="SPTREMBL:Q9JR17"
/translation="MKDKHSSAMRLDKLWAAFFKTRSLAQKHIELGRVQVNGSKV
DNKSTIDIGDIIDLTLNSLPYKIKVGLNHNORRPAEARLLYEADKATLREERKOL
QFSRTTSAYPDGRPTKRRRLDKLKKGM"
complement(247675..247860)
/gene="NMA1348"
/note="Pfam match to entry PF01479 S4, S4 domain, score
59.00, E-value 1e-13"
248099..249058
/gene="acca"
248099..249058
/gene="acca"
/EC_number="6.4.1.2"
/note="NMA1349, acca, probable acetyl-CoA carboxylase
carboxyl transferase subunit, len: 319 aa; similar to many
e.g. SW:ACCA_ECOLI (EMBL:M96394), acca, Escherichia coli
acetyl-CoA carboxylase carboxyl transferase subunit (EC
6.4.1.2) (318 aa), fasta scores: E(): 0, 64.9% identity in
313 aa overlap"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase carboxyl
transferase subunit"
/protein_id="CAB84596.1"
/db_xref="GI:7380018"
/db_xref="SPTREMBL:Q9JUF0"
/translation="MKPVFLDFEQPTAELTNKIDELRFVQDESVIDSDETHRLQKKS
NDLTKSIFSKLTAPQISOVSRHPORPYTLDY IEALFTDFEELHGDGRFADDDYAEVGL
AFENGQVMVVGHQKGRDTEKELIRNFGMPRECYRKALRLMKTAERFGLPVWTFIDT
PGAYPGIGAERQCSAIGKALNELFLRLRPVPLCTVIGCGSGGALAVAGVDYNNMLQ
YSTSVISPECCASILMKTAEKADAQAUGITADRLQKLDLVDTVITKEPLGGAHROF
GQTMKNVAYLEKQLHEAQSIPLADLULSRFRDIMAYGKFSEQ"
249156..250436
/gene="NMA1350"
249156..250436

```

AAGATVILVARHOKKLEKVYDAIVEARHPPEFAIRFDLMSAEKEEFNOFAATIAEATQ
GKLDGI1VHCAGYFYALSPDLDFQTAEWNOYRINTVAPMGLTRALFPLLLKQSPDASVI
FVGSHEGTPKAYWGFGCASKAALNYLCKVAADWEERFGLNRANVLVPGPINSQRIK
SHPGEAKSRKSYGDLVLPFAVFWMASTESKGRSGEIVL"
complement(239082..239654)
/gene="NMA1336"
/notes="Pfam match to entry PF00106 adh_short, short chain
dehydrogenase, score 130.20, E-value 3.9e-35"
complement(239755..239764)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(239756..241069)
/gene="NMA1338"
complement(239756..241069)
/notes="NMA1338"
/notes="NMA1338, possible oxidoreductase, len: 437 aa;
shows weak similarity to phytoene dehydrogenases e.g.
TR:Q40406 (EMBL:X78815), pds, Narcissus pseudonarcissus
phytoene dehydrogenase precursor (570 aa), fasta scores;
E(): 1.6e-06, 24.7% identity in 473 aa overlap. Also to
SW:Y4AB_RHISN (EMBL:AE000064), Y4AB, Rhizobium sp. (strain
NGR234) hypothetical protein (417 aa), fasta scores; E():
8.9e-15, 27.3% identity in 444 aa overlap"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAB84587.1"
/db_xref="GI:7380009"
/db_xref="SPTREMBL:Q9JUF6"
/translation="MMNTPHPRKIAVIGAGWAGLSAAVTLARHADVTLPFAGRQAGG
RARALAGNTDGFGLDNGQHILGAYRGVLRMLKTIQSDPHAAFLRPLHWHMHGGLQ
FRALPAPLHILGVLRLARRVPSAFKALLADMQLDQKLSRGSDPTTVAOMLKORN
VPRAAVMOFQIPLWGLNLTPEASLRVLCNVLSQGLTKKSGSVLLPKQDLGAIY
AEPALAEQNLRGADIRLETRICRLNLPDGKVLVNGPEPDAAVPATAPYHAAALLPEG
TPHVOTATONLYRHAITYLVYLRAPVRLPAPLGLADGTVOMLLCRGLGUPENEV
SAVISVSDRVGAFANRAWADKRVHADLRILPHIGEPEAVRVIETKRATTAADAPPPDL
SWLHRIRIFPAGDYLHPDYPATLEAAVQSGFASAEACLOSLSDAV"
240219..240228
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(240342..240351)
/gene="NMA1338"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(241075..241079)
complement(241150..242022)
/gene="NMA1339"
complement(241150..242022)
/gene="NMA1339"
/EC_number="2.5.1.-"
/notes="NMA1339, possible poly-isoprenyl transferase, len:
290 aa; similar to many phytoene synthases e.g.
SW:PSY_CAPAN (EMBL:X68017), psy1, Capsicum annuum phytoene
synthase precursor (EC 2.5.1.-) (419 aa), fasta scores;
E(): 6.7e-26, 36.0% identity in 278 aa overlap (NMA1339 is
shorter at the N-terminus) and TR:CAB39693
(EMBL:AL049485), SC6A5.09, Streptomyces coelicolor
putative phytoene synthase (312 aa0, fasta scores; E():
1.6e-32, 38.0% identity in 287 aa overlap. Similar to
NMA1721, fasta scores; E(): 2e-17, 31.0% identity in 226
aa overlap. Contains Pfam match to entry PF00494 SQS_PSY,
Squalene and phytoene synthases"
/codon_start=1
/transl_table=11
/product="putative poly-isoprenyl transferase"
/protein_id="CAB84588.1"
/db_xref="GI:7380010"
/db_xref="SPTREMBL:Q9JUF5"
/translation="NKGDLICROKAEBSRSLSGFRFLTOEKRDVATVLYAFCEFLD
DYVDCSNPDVAQATLNNWRGDLKVFVGGAPEHVNQALROYKETFPLPKYELKALI
DGMQMDLLRVYGSPEELKYCRRVAGVVGCLIAKILGFSDDQTLLEYADKMGLALQLT
NIIRVDGARGRIYLPMEEREDVPASVILQCSPTGNFAELMAFIKRAETRE
AVSLLPADAKKQAKVGLVMAAVYYELLNEIDRDQGNVLKYKIALPSRKKRIALTKT

```

```

LFGPKPRPGCTPERA"
complement(241366..241995)
/gene="NMA1339"
/notes="Pfam match to entry PF00494 SQS_PSY, Squalene and
phytoene synthases, score 198.00, E-value 1.4e-55"
complement(242032..242035)
242193..244055
/gene="NMA1340"
242193..244055
/gene="NMA1340"
/notes="NMA1340, probable chaperone protein, len: 620 aa;
similar to many e.g. SW:HSCA_ECOLI (EMBL:U01827), hscA,
Escherichia coli chaperone protein (616 aa), fasta scores;
E(): 0, 57.8% identity in 626 aa overlap. Similar to
NMA0736, fasta scores; E(): 0, 43.0% identity in 596 aa
overlap. Contains Pfam match to entry PF00012 HSP70, Hsp70
protein"
/codon_start=1
/transl_table=11
/product="putative chaperone protein"
/protein_id="CAB84589.1"
/db_xref="GI:7380011"
/db_xref="SPTREMBL:Q9JUF4"
/translation="MALQISEPGMSAAPHRHRLAAGIDLTNTNSLVATVRSQAACL
PDAGRVTLFSVRYLENGGIEVGTALSAQKTDPLNTVSSAKRLIGRTLADLHQNTH
LYLPRFGDNQRIELHTRQGVKTPVEVSAETILTKLSRAEETLGGDLGVGVITVPAYF
DDAQKATKDAARLAGLNLRLLENEPTAAATAYGLDNASEGTFVYDLGGTFFDVSVL
LTKGFEVKGATGNSALGDDFDHRLFCRLLEQNGLSQLEODSOLLKSLVRAAKEQ
LTQTEARIQATLSDGMPIDTSISRAEFHNLTOHLVMKLTLEPVTQALKDAGVCKNEVK
GLVMGSGTRMLHVOQAVTFEGQTPLNLPNPEVVALGAAIQANVLAKNKTGDEWLL
LDVTPLSLGETVGGLAEKIIPRNSTIPYARQDFTTFKQGTQAMTHVVOGERELVS
DCSLAKFTLURGIPMAAGAAARIVTQIDAGLLSSAQEQSGTQVQAQIEVPSYGL
DGAITRMLKDSMDNAAEAAARAAYVVEESLTDVNAALFLDSDLDLDAKELQOI
ROGIADLQGRLLKDGKABEDIRSAVAKLSRSTDNFAAKRMNRNIQKALTQGSVDNI"
242253..244007
/gene="NMA1340"
/notes="Pfam match to entry PF00012 HSP70, Hsp70 protein,
score 767.70, E-value 4.6e-227"
complement(243059..243068)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
243916..243925
/gene="NMA1340"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(244068..244077)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
244094..244103
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(244183..244337)
/notes="Correia element; hmms hit to HMM Correia (1
156), score: 251.87"
/label=Correia
244378..244941
/gene="NMA1341"
244378..244941
/gene="NMA1341"
/notes="NMA1341, len: 187 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1341"
/protein_id="CAB84590.1"
/db_xref="GI:7380012"
/db_xref="SPTREMBL:Q9JUF3"
/translation="MRPLENNYANVAGLERLAVKTDYKKNELLHEIFSKSRIGD
TELFPAVDENLVKRLFLSLRGEIVFPKANTESEFEKSVHRRQEGNAGSKQLDLV
RGHREYVPLRPLAGAAAYPKPKSKIRLFEAVFGKSGTRLTDTDIADGIIHYTCFS
RADLKAYSEYLELKFESDAEGRKPQ"
245178..245738
/gene="NMA1343"
245178..245738
/gene="NMA1343"
245178..245738

```

```
/db_xref="GI:7380002"
/db_xref="SPTREMBL:Q9JUG0"
/translation="WTIFYKNGFYDDTLGSIPEGAVAVRAEYEAALLAGAOGOIAA
DSDGRVLPPRPSEYHEWDGKKEWGEARAAAEQKTATAFRIAARADELKNLL
AGYPOVEIDFYRQEKALRAQADNNAPTPMLAQTAARGVELDVLLIEKVVESKARLA
VAAGAIIGRQOLEDKLNTIETAPGLDALEKEETEWTLIG"
RBS
234886. .234889
/gene="NMA1328"
234899. .235195
/gene="NMA1329"
234899. .235195
/gene="NMA1329"
/notes="NMA1329, len: 98 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1329"
/protein_id="CAB84580.1"
/db_xref="GI:7380003"
/db_xref="SPTREMBL:Q9JUF9"
/translation="MDAKHRLKXYVHLLVAIDQLFNALTGGADETLSRTRYGARL
AOKPKTRWKVLYTLNGVEFDHQHCRQAVISELKGHQHARENQSRAGEKQTR"
235192. .235401
/gene="NMA1330"
235192. .235401
/gene="NMA1330"
/notes="NMA1330, len: 69 aa; unknown, similar to the
C-terminus of SW:YF23_HAEIN (EMBL:U32828), H11523,
Haemophilus influenzae hypothetical protein (296 aa),
fasta scores: E(): 2.3e-07, 33.9% identity in 59 aa
overlap and to the C-terminus of NMA1821, fasta scores;
E(): 1.7e-07, 39.3% identity in 56 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1330"
/protein_id="CAB84581.1"
/db_xref="GI:7380004"
/db_xref="SPTREMBL:Q9JUF8"
/translation="MNYFGMVEFLMAMVRPFVFFSGTRSELPYLLDLVAELRTG
WERFAGSQLTVYSTSTAIAQT"
235511. .236653
/gene="NMA1331"
235511. .236653
/gene="NMA1331"
/notes="NMA1331, pseudogene, ABC-transporter ATP-binding
protein, len: 1140 bp; similar to the N-terminal half of
many putative ABC-transporter ATP-binding proteins e.g.
TR:O34061 (EMBL:AJ225561), exsE, Sinorhizobium meliloti
putative ABC-transporter ATP-binding protein (606 aa),
fasta scores: E(): 4.1e-20, 27.7% identity in 321 aa
overlap and SW:YE67_HAEIN (EMBL:U32825), H11467,
Haemophilus influenzae putative ABC-transporter
ATP-binding protein (589 aa), fasta scores; E(): 0, 49.0%
identity in 355 aa overlap Note that the C-terminal half
of this protein appears to be NMA1280, beyond the
intervening prophage. Contains hydrophobic, possible
membrane-spanning regions"
/codon_start=1
/pseudo
/transl_table=11
/product="ABC-transporter ATP-binding protein
(pseudogene)"
235583. .235592
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/misc_feature
/label=DUS
236677. .236686
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(236760. .236914)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 277.83"
/misc_feature
/label=Correia
237058. .237067
/notes="Core DNA uptake sequence: gcgcgtctgaa"
```

```
/label=DUS
complement(237095. .237742)
/gene="NMA1332"
complement(237095. .237742)
/notes="NMA1332"
/gene="NMA1332, possible periplasmic protein, len: 215 aa;
unknown, contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84583.1"
/db_xref="GI:7380005"
/db_xref="SPTREMBL:Q9JQV3"
/translation="MKPLILGLAAVLASACQVQKAPDFDYSFKESKPASILVVPPL
NESPQVNGTGWGLASTAAPLSEAGYVFAAVVEFKQNGLTNAADIAHAREKELHQ
IFGNDALVITVTEYGTSYQILDSTVTVSAKARLVDSRNGKELWSSASITREGSNNSN
SGLLGALVASVNVQIANSLTRGYSQVSKTAAYNLLSPSYSHNGILKGPREVEQPK"
complement(237739. .238110)
/gene="NMA1334"
complement(237739. .238110)
/notes="NMA1334"
/notes="NMA1334, possible periplasmic protein, len: 123 aa;
unknown, contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84584.1"
/db_xref="GI:7380006"
/db_xref="SPTREMBL:Q9JQZ2"
/translation="MMNPRTLRLSLCAAVALTACGGNGQKSLYYGYGYPDTVYBGL
KNDDTSLGQTEKMEKYFVEAGNKKMNAAPGAHAHLGLLSRSGDKGEGAFRQPEEKR
LPESGVPMDFLMKTKGKGR"
complement(237749. .237754)
/gene="NMA1334"
complement(238113. .238117)
complement(238117. .238788)
/gene="NMA1335"
complement(238117. .238788)
/notes="NMA1335"
/notes="NMA1335, possible periplasmic protein, len: 223 aa;
unknown, contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84585.1"
/db_xref="GI:7380007"
/db_xref="SPTREMBL:Q9JRF1"
/translation="MKTVSTAVVLAAAVSLTGCAATESRSLEVEKVASINTQYHGVR
TPISVGTDFNRSFQKGFISDGEDRLGSOAKTILVHLQQTNRFNVLNRTNLALQKE
SGISGKAHNLKGADYVVTGDVTEFGRRDYGHQILGILGRKSOIAYAKVALNIVNVN
TSEIYVYSAQGAGEYALSNREIIGFGTSGYDATLNGKVLDLDAIREAVNSILVQAVDNGA
WQPNR"
complement(238953. .239723)
/gene="NMA1336"
complement(238953. .239723)
/notes="NMA1336, possible oxidoreductase, len: 256 aa;
similar to many from both bacterial and eukaryotic sources
e.g. weak similarity to SW:TRN1_DAYST (EMBL:L20473) Datura
stramonium tropinone reductase-I (273 aa), fasta scores;
E(): 7.1e-11, 30.2% identity in 205 aa overlap. Similar to
SW:YCIK_ECOLI (EMBL:AE000225), yciK, Escherichia coli
hypothetical oxidoreductase (252 aa), fasta scores; E():
6.9e-25, 35.2% identity in 233 aa overlap. Contains Pfam
match to entry PF00106 adh_short, short chain
dehydrogenase"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAB84586.1"
/db_xref="GI:7380008"
/db_xref="SPTREMBL:Q9JUF7"
/translation="MRQNAKIPPTDQKKGKSMATLSDKTILITGASOGLGEQVAKAYA"
```

```

/db_xref="SPTREMBL:Q9JUG6"
/translation="MDKELNFGTGYDTGRTVDTLQNAVYIRLMTPLGSGWADKTLGSL
LHLQREKDIQVSLAEQYADEALQPIVSGRADKITVRAEQPHDGRLLIHIRVDTA
AGGFDYRHEVPVI"
229628..229632
229641..230696
/genes="NMA1323"
229641..230696
/genes="NMA1323"
/notes="NMA1323, len: 351 aa; unknown, similar to bacterial
hypothetical proteins including SW:KXDT_BACSV
(EMBL:270177), xkdt, phage-like element Pbsx protein (348
aa), fasta scores; E(): 1.2e-07, 28.0% identity in 361 aa
overlap and SW:YE20_HAEIN (EMBL:U32827), H11520,
Haemophilus influenzae hypothetical protein (355 aa),
fasta scores; E(): 4.7e-16, 31.5% identity in 321 aa,
overlap. Similar at the N-terminus to NMA1827, fasta
scores; E(): 1.5e-18, 41.4% identity in 152 aa overlap and
at the C-terminus to NMA1826, fasta scores; E(): 1e-29,
50.8% identity in 185 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1323"
/protein_id="CAB84574.1"
/db_xref="GI:7379997"
/db_xref="SPTREMBL:Q9JUG5"
/translation="MFETPTFEQIRERILRDTKSLWPDADISPDSDHYVHASRLASCA
EGOYAHQSWIVROIFPDADREYLERHSMRGLRRNPPTASGTLVSGTAQSLSDG
LVYIGORFYRTTARAVIGSGTAETIPADEPAAANVRDGEAQLMAAPAGVSTECR
LTVOGQTDRESASLARLEIIRPPAGNRYDYKNWALSVDGVTSAVYVPLRLGIG
TVDIATSDGVPSEETVRRVQAYIDEMRPVTAKNALVKPTVAVPTVQVKLDGID
LDEAKRRIETALKEFDLIPGDGLTVSQIEAASISNVGVIDLRLTAPTANRAADTVN
RIEMFKAGAINVTMPDS"
complement(229937..229946)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
230681..230684
/genes="NMA1323"
230693..231253
/genes="NMA1324"
230693..231253
/genes="NMA1324"
/notes="NMA1324, len: 186 aa; unknown, shows weak
similarity to SW:YMFQ_ECOLI (EMBL:AB000214), ymfQ,
Escherichia coli hypothetical protein (194 aa), fasta
scores; E(): 0.0001, 28.2% identity in 195 aa overlap.
Similar to NMA1825, fasta scores; E(): 9.6e-24, 42.4%
identity in 184 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1324"
/protein_id="CAB84575.1"
/db_xref="GI:7379998"
/db_xref="SPTREMBL:Q9JUG4"
/translation="MSYDILKGLLPVSVYARNAPRVRAQAEIDGAALDAVAESAQSV
ADAVPSGQALMDWVERVLGDGTCKNRRVLVMAKLNETGISPIFYRLAEAA
GYQIQTDEQPPFRAGVNRAGDLAPQEIIMVWVHVNRGCGNNRITFRAGISAAGDLT
DYSDAVIESLNRKLPAPHTAIRFTYR"
231253..231256
231264..233237
/genes="NMA1325"
231264..233237
/genes="NMA1325"
/notes="NMA1325, possible phage tail fibre protein, len:
657 aa; similar in part to SW:VPH_BPHPI (EMBL:U24159)
bacteriophage Hp1 probable tail fibre protein (925 aa),
fasta scores; E(): 0, 48.7% identity in 452 aa overlap.
The N-terminus is similar to that of NMA1824, fasta
scores; E(): 3.9e-09, 33.3% identity in 147 aa overlap"
/codon_start=1
/transl_table=11
/product="putative phage tail fibre protein"
/protein_id="CAB84576.1"

```

```

/db_xref="GI:7379999"
/db_xref="SPTREMBL:Q9JUG3"
/translation="MHPJETPKTFHGDGVSELGITLPAWMLNQVSELAVILTAAG
IQPKSOPNOLLAALNRLLVVVITANQETFTAQTFQSGIHLSANQTNWNGGHK
AYIGADADNAHIVFGDDTLRLHSANNRISYNNHIDIFHKANKPRAEDIEGKPNLTLSGY
GIGNFKVETFRGDLNLTKTGDIYSLPTAVSGSNLPIVENTACHIOVIAGTKHWCRCOLG
YPAYTSDVYERHOTSSANDNSAMKKGNDGIPVGAIVSPFKAVRNPAGYILKANGTTF
AONTFDPDLYRALGNSNRLPDLSTRDIGITANFPDQIPTGLAFDDIIRTVTKTAIYPE
LRYLLTKGYSIQNVPQAREFRIRNAANGLAVGTKQDEDEIKRHVKHVFHSWNNHPHAA
ALGYEDHNEROSALSALVSTWTDENLNDGFLTPRPSDKMATGGDENRKRKALVILKICIA
ADTLEAVFWIKSHGETVAGALDAGTIAQLQDKADRAHTTAAQIOGLQDEKISTVSE
AAQFTROTITIGVDIVRFPGDTMIOTGSYRFRSGGPIENEVVFYVADAGNVKCFVSE
RHSGRVTTGDRQHNWLFIRAKNHAALITNNYEGSCDWMALGKASGNAASSPIGPPEI
PETNEEQRESGRSTGPRNRHRDLSLEALQD"
233387..233406
/notes=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(233407..233462)
/notes="83% identity to consensus
CCAGTCCGTTTCAGTTTCAGTCATTCGCATAAATTCCTGTGCTTTTCATTCTTAG"
/label=RS17
233463..233482
/notes=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
233504..233523
/notes=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(233524..233542)
/notes=">= 85% match to ATCCCNACNACNTTTCGTC"
/label=RS3x
complement(233674..233976)
/genes="NMA1326"
complement(233674..233976)
/genes="NMA1326"
/notes="NMA1326, len: 100 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1326"
/protein_id="CAB84577.1"
/db_xref="GI:7380000"
/db_xref="SPTREMBL:Q9JUG2"
/translation="MOROSLTNADGEVRELAEDFALARDIAELPDALQVLFESHOK
OLEEKGVMOKRNTGKTPKQOLVTIRLSADVVEKFRAGKGWQTRINEVIRQYVQLK"
complement(233948..234235)
/genes="NMA1327"
complement(233948..234235)
/genes="NMA1327"
/notes="NMA1327, len: 95 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1327"
/protein_id="CAB84578.1"
/db_xref="GI:7380001"
/db_xref="SPTREMBL:Q9JUG1"
/translation="MKIEFDSKQNRNIBERNLPFESVGQIRHTTAVIVDPVCFDYPE
PRVYAAAYLGDQRLHIVCFITGIDGIRVISFRKANKREVKKYATYSILNKR"
complement(233986..233990)
/genes="NMA1327"
234302..234305
234315..234920
/genes="NMA1328"
234315..234920
/genes="NMA1328"
/notes="NMA1328, len: 201 aa; unknown, C-terminal half is
similar to SW:YE02_HAEIN (EMBL:U32820), H11402,
Haemophilus influenzae hypothetical protein (68 aa), fasta
scores; E(): 1.3e-08, 56.3% identity in 64 aa overlap.
Shows weak similarity to NMA1823, E(): 3.6e-07, 28.8%
identity in 191 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1328"
/protein_id="CAB84579.1"

```

Haemophilus influenzae hypothetical protein (169 aa),
fasta scores; E(): 4.6e-25, 48.7% identity in 154 aa
overlap. SW:YF02_HAEIN (EMBL:U32827), H11502, Haemophilus
influenzae hypothetical protein (414 aa), fasta scores;
E(): 3.8e-22, 35.4% identity in 237 aa overlap and
TR:Q46548 (EMBL:U20249) Bacteroides nodosus plasmid
pJIR787 ORF240* (fragment) (240 aa), fasta scores; E():
4.8e-12, 30.4% identity in 194 aa overlap. Similar to
NMA1850, fasta scores; E(): 0, 33.1% identity in 378 aa
overlap".
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1315"
/protein_id="CAB84567.1"
/db_xref="GI:7379991"
/db_xref="SPTREMBL:Q9JUHO"
/translation="MPADPLFALSLPPKKAIEWLKSKVTAESVRNLTASEIAKYVT
IARTDMLNDIKTSMVESAKSGQFDDWRKGLNLLSNKGLHPNGHNGKDIIDPA
TGEVFGSPRRLEIITYTNNGYACNAGQYQGMANIDARPYWMDVAGDSRTRPAHSAI
DGLVRYDDPTWATYTPNGYNCRCVIALSERDVERGRIVGQSTSNLVETHKIYN
KKGDTYLTAVKAPDGSLYTTDRGFDYNAGRMNRPDLDKYDRALAHOPAKAEMGGAD
FKTSEKQLEFEYEVKQRLDIDGRFDKEQIKIRNALSRQLKFAAGVLSKETQELAGM
TRATVWLSDDLVVKQVDSREGQFDDSYAFPLDMLQNPHEVIRNRELIFTARYKGS
ALWAVLYKIEVDEIYQSYRISNDKETAKFMKKVLK"
/gene="NMA1316"
/note="NMA1316"
/db_xref="SPTREMBL:Q9JUHO"
/translation="MPADPLFALSLPPKKAIEWLKSKVTAESVRNLTASEIAKYVT
IARTDMLNDIKTSMVESAKSGQFDDWRKGLNLLSNKGLHPNGHNGKDIIDPA
TGEVFGSPRRLEIITYTNNGYACNAGQYQGMANIDARPYWMDVAGDSRTRPAHSAI
DGLVRYDDPTWATYTPNGYNCRCVIALSERDVERGRIVGQSTSNLVETHKIYN
KKGDTYLTAVKAPDGSLYTTDRGFDYNAGRMNRPDLDKYDRALAHOPAKAEMGGAD
FKTSEKQLEFEYEVKQRLDIDGRFDKEQIKIRNALSRQLKFAAGVLSKETQELAGM
TRATVWLSDDLVVKQVDSREGQFDDSYAFPLDMLQNPHEVIRNRELIFTARYKGS
ALWAVLYKIEVDEIYQSYRISNDKETAKFMKKVLK"
/gene="NMA1316"
/note="NMA1316, len: 138 aa; unknown, shows weak
similarity to SW:VPG_BPMU (EMBL:M74911), bacteriophage Mu
G protein (156 aa), fasta scores; E(): 0.0065, 29.9%
identity in 87 aa overlap and SW:VPG2_HAEIN (EMBL:U32831),
H11568, Haemophilus influenzae putative cryptic M-
protein 2 (138 aa), fasta scores; E(): 1.3e-22, 47.4%
identity in 137 aa overlap. Similar to NMA1849, E():
1.6e-11, 38.7% identity in 163 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1316"
/protein_id="CAB84568.1"
/db_xref="GI:7379992"
/db_xref="SPTREMBL:Q9JUG9"
/translation="MIDVKIDNIFVLNQIRLGNIGIENRLLMRLSETMHTAVKLN
FRYGRPKWLGLKYRDKPLSDSGRLKDSFTLSNDTALVGTNIVYAAIHFGMGAG
RNRKVRIPQRELTLTDDDKQALMDVDQDYSGLIP"
/gene="NMA1318"
/note="NMA1318, pseudogene, possible phage protein, len:
1059 bp; similar to SW:VPI_BPMU (EMBL:M74911)
bacteriophage Mu I protein (fragment) (58 aa), fasta
scores; E(): 0.12, 51.5% identity in 33 aa overlap
SW:VPI_HAEIN (EMBL:U32827), H11504, Haemophilus influenzae
putative cryptic Mu-phage I protein (355 aa), fasta
scores; E(): 5.8e-14, 31.6% identity in 269 aa overlap.
Similar to NMA1848, fasta scores; E(): 2e-19, 39.1%
identity in 261 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative phage protein (pseudogene)"
/gene="NMA1319"
/note="NMA1319, len: 156 aa; unknown, similar to part of
NMA1831, fasta scores; E(): 6.4e-12, 46.3% identity in 95
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1319"

/protein_id="CAB84570.1"
/db_xref="GI:7379993"
/db_xref="SPTREMBL:Q9JUG8"
/translation="MQPVAQAVRLLSTSSLLSVATALIEAHGEEMTAPDLIEVNRAMR
PLMQEIAALRAVQTAAAEESGLTANAVYTEAYQTAEASLRAGAARUNALVAAVINOKP
RMVROAPRIDGTIHQIAHEFYGDIAAAELVRLNPHIHPAFIKRGTLVNSYAK"
227172. .227181
/gene="NMA1319"
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
227370. .228515
/gene="NMA1320"
227370. .228515
/gene="NMA1320"
/note="NMA1320, len: 381 aa; unknown, shows weak
similarity to SW:VPP_BPMU (EMBL:X06796) bacteriophage Mu
tail protein (379 aa), fasta scores; E(): 1.2e-12, 24.9%
identity in 366 aa overlap. Similar to NMA1830, fasta
scores; E(): 0, 44.8% identity in 373 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1320"
/protein_id="CAB84571.1"
/db_xref="GI:7379994"
/db_xref="SPTREMBL:Q9JUG7"
/translation="MNNNSYGVAVSVRYGKREHWRVYDIDSLIPADSDFVIGR
LGPEAAIPDLSECEVVDIGQIVMTGLIGSORHGKSGKRELSLSGRDLAGFLVDCS
APQLNVMGMTVDAAKLAAPPOKIAVVKVNNPALDKDIEGETVWQALTHIAN
SVGLHPLEPDGTLVGVYSSPPVATLCWSRTDSRRNIEMDJEWDTDNRFSEVTF
LAQSHRSGDSAKHDLKWYKDPMTLHRPKTVVYSDADNLAALQAKKQADWRLE
GKTITVTGGHKTDRGVLMQQRVHVDDHGHGIDAVFLMGRRLMRLSMDGTQTELR
LKEDGIWTPDAYPKKAEAAARKRKRGVSHKGGKGGKQAEATVFE"
228512. .229180
/gene="NMA1321"
228512. .229180
/gene="NMA1321"
/note="NMA1321, len: 222 aa; unknown, shows weak
similarity to SW:VPV_BPP2 (EMBL:AF063097) bacteriophage P2
baseplate assembly protein V (211 aa), fasta scores; E():
0.85, 24.4% identity in 221 aa overlap and SW:YF18_HAEIN
(EMBL:U32827), H11518, Haemophilus influenzae hypothetical
protein (182 aa), fasta scores; E(): 0.0011, 26.8%
identity in 194 aa overlap. Similar to NMA1829, fasta
scores; E(): 1.8e-32, 46.7% identity in 197 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1321"
/protein_id="CAB84572.1"
/db_xref="GI:7379995"
/db_xref="SPTREMBL:Q9JQO9"
/translation="MSLSKLAKKTAOTAKNIGETLRAAFRGKITLVVSEPIQRYOLS
GLADETLDLHLEQYGFASHPDGSSEAVVPLGNTSHGVIVCSQHSYRIKLNKPG
ETAIFNHEGAKIVIKQGIIEADCDVYVNCQYEVNAATAKFNAPLIEVAVLTAQ
GOINGNGMVEGGDGAFTSGDVNQTGGSFNTDGDVAVAGNLSLRQHPHTDSIGGKTLP
AEP"
229253. .229262
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
229280. .229627
/gene="NMA1322"
229280. .229627
/gene="NMA1322"
/note="NMA1322, len: 115 aa; unknown, shows weak
similarity to SW:YF19_HAEIN (EMBL:U32827), H11519,
Haemophilus influenzae hypothetical protein (135 aa),
fasta scores; E(): 0.031, 36.3% identity in 102 aa
overlap. Similar to NMA1828, fasta scores; E(): 3.4e-20,
51.3% identity in 117 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1322"
/protein_id="CAB84573.1"
/db_xref="GI:7379996"

```

/product="hypothetical protein NMA1308"
/protein_id="CAB84560.1"
/db_xref="GI:7379984"
/db_xref="SPTREMBL:Q9JUH7"
/translation="MTTYRELQVTRACRHADLELGLSRAREQPFVHVSDLLDKAG
IDYAVRMKDFQTFVCVEFSATPLLM"
219399..219518
/gene="NMA1309"
CDS
219399..219518
/gene="NMA1309"
/notes="NMA1309, len: 39 aa; unknown, similar to NMA1195,
fasta scores; E(): 6.1e-08, 65.7% identity in 35 aa
overlap and NMA1857, fasta scores; E(): 0.036, 42.9%
identity in 35 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1309"
/protein_id="CAB84561.1"
/db_xref="GI:7379985"
/db_xref="SPTREMBL:Q9JUH6"
/translation="MIGILRKYYVFSFGQKVEASRHPGYAVRIVFGDVPV"
complement(219434..219443)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
219530..219796
/gene="NMA1310"
CDS
219530..219796
/gene="NMA1310"
/notes="NMA1310, len: 88 aa; unknown, highly similar to
NMA1196, fasta scores; E(): 3.9e-33, 94.3% identity in 88
aa overlap and similar to NMA1156, fasta scores; E():
1.5e-10, 44.0% identity in 84 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1310"
/protein_id="CAB84562.1"
/db_xref="GI:7379986"
/db_xref="SPTREMBL:Q9JUH5"
/translation="MDFERGFKTLWPIATAAFWFWNGISGRLEKADKRIDDLKEELH
AVKLSYHTQDAKADSTNIAALERIENLEKVNELDRKADKS"
219609..219618
/gene="NMA1310"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
219793..219990
/gene="NMA1311"
CDS
219793..219990
/gene="NMA1311"
/notes="NMA1311, len: 65 aa; unknown, similar to NMA1197,
fasta scores; E(): 1.1e-14, 56.3% identity in 64 aa
overlap and NMA1855, fasta scores; E(): 1.2e-10, 47.0%
identity in 66 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1311"
/protein_id="CAB84563.1"
/db_xref="GI:7379987"
/db_xref="SPTREMBL:Q9JUH4"
/translation="MSDPILEVIARIEAKQDDMLANQARMDELOQIKKDKCKSAAYV
GGIGGVIVTTGWELLRAKFGG"
219992..220498
/gene="NMA1312"
CDS
219992..220498
/gene="NMA1312"
/notes="NMA1312, possible DNA-binding protein, len: 168 aa;
unknown, similar to NMA1198, fasta scores; E(): 0, 73.7%
identity in 167 aa overlap and NMA1854, fasta scores; E():
0, 56.3% identity in 167 aa overlap. Contains probable
helix-turn-helix motif at aa 21-42 (Score 1051, +2.77 SD)"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="CAB84564.1"

```

```

/db_xref="GI:7379988"
/db_xref="SPTREMBL:Q9JUH3"
/translation="MAHPKEIREKRLRLLYVSDGOTLEIAAMCMCIPIATARSWKRAAK
ESSDDWKVRAAYTLAGGIEDLSRLSLAGFLVOYQSTMTLMQDTSLEGLMPSERAKL
LTSLSDAFTKTVAAANKVMPETSKLATAIEVLELFGVWRERYPQHLOAQYVELVEPLG
VETEKKYR"
220519..222141
/gene="NMA1313"
CDS
220519..222141
/gene="NMA1313"
/notes="NMA1313, len: 540 aa; unknown, highly similar to
NMA1852, fasta scores; E(): 0, 86.5% identity in 539 aa
overlap. Contains PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1313"
/protein_id="CAB84565.1"
/db_xref="GI:7379989"
/db_xref="SPTREMBL:Q9JUH2"
/translation="MKSKEFLKSLAEYAAQLRIIEAEVDGFDASPKAIAARRAKVFD
PVGGYEYFNTYFPHYIRSEKSELHAFLSRLPEIIRSPKGENEAVGAPRGEGKSTQ
VTQLFTLCIVTGOKHYAVIVMDSIDQAYPMLEAKAELEFNRLKYDFEVCQGRV
WQAGTIVTANDVKOVAGSGKLRGLRHGYPYRDLTVLDDIENDEQVRNPEQRDKLNA
WLTKYVLPGLGVGQKDYIYITILHYDSVLNLTNNPFWHGKFKKAMKMPDRMDLW
DRWEELFRNGDETVAEAFILANKDEMERGAQTSWAARGVLALMKIRADGHATFDESY
ONDPVSGEDAPFAKSMKFWNDLPSDLVYFGALDPSLGAGASRDPISAIVIGGYQSMG
RLYVVEAQIKKRLPDLIEDVICLHRQYRCKLMFEVTVQFEFLKDELVKRSAARGIP
PVARAKPVSKLLRIETLQHPMANGILLINESQOTLIQFRHFPKADHDGDPDAVHM
LWSGAVANCYPIEQWSPTDNDFDEIKKWSR"
220795..220818
/gene="NMA1313"
misc_feature
220795..220818
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
222131..222134
/gene="NMA1313"
RBS
222141..223709
/gene="NMA1314"
CDS
222141..223709
/gene="NMA1314"
/notes="NMA1314, len: 522 aa; unknown, similar to
SW:YF01_HAEIN (EMBL:U32826), H11501, Haemophilus
influenzae hypothetical protein (520 aa), fasta scores;
E(): 0, 40.2% identity in 518 aa overlap and contains a
region similar to SW:YF71_HAEIN (EMBL:U32831), H11571,
Haemophilus influenzae hypothetical protein (79 aa), fasta
scores; E(): 7.5e-14, 63.2% identity in 76 aa overlap.
Highly similar to NMA1851, fasta scores; E(): 0, 49.0%
identity in 510 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1314"
/protein_id="CAB84566.1"
/db_xref="GI:7379990"
/db_xref="SPTREMBL:Q9JUH1"
/translation="MAKNKNTKIQPEAAALQTDVAQITATGRVIAHPNSFTTPQKM
RALFDEASGDIRAQHELFAIEERSDIAANNTRKALLTLNWRVAPPRNATPEEE
KLSDQAYEMMDSLTLEDLMDLDAVGHGFSALEVEVSDGLYLPNRIHPRQSWF
KWDKDNGLLIRTRNPEGEALPLGWVYVTKSRVQOARNGFLRTLISWLYMFKHVAV
HDPFAEFLYGMPIRIGKYGAGATKEEKNTLLRAVAEIGHNAGIMPEGHEIELHNA
NGMTSAGNPLQWADNCEKSAARLIGQTLTSGDGSKSNALGNHNEIRDLVSD
AKGVAQITISQIIGPELQINYPHADPNRPVKFEFTREPDKIAFADAIKPLKVDVGQ
IPESWYRKLVIPDQGEAVLVYRQVDPNPNRTALAAISAHTVPVSKATGRHQEILDG
ALDDALVEPDFNSQLNPMVRQVAALNACNSYEADAALNALYNLNDNAKLRTYMQQA
LFTSDILGDHARA"
complement(222566..222575)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
223696..224991
/gene="NMA1315"
CDS
223696..224991
/gene="NMA1315"
/notes="NMA1315, len: 431 aa; unknown, contains regions
similar to SW:YF70_HAEIN (EMBL:U32831), H11570,

```

```
misc_feature
/label=DUS
complement(216163..216172)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
216223..216717
/genes="NMA1301"
216223..216717
/genes="NMA1301"
/note="NMA1301, len: 164 aa; unknown, similar to part of
TR:Q38494 (EMBL:M64097) bacteriophage Mu protein E16 (195
aa), fasta scores; E(): 2.7e-05, 46.7% identity in 60 aa
overlap. Identical to NMA1186 and similar to NMA1867,
fasta scores; E(): 1.5e-25, 55.5% identity in 137 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1301"
/protein_id="CAB84553.1"
/db_xref="GI:7379977"
/db_xref="SPTREMBL:Q9JS31"
/translation="MYFIGICVNSIYCVYEFEMRRALIAKIAQKELGDDGTyr
AVLERVTDKRSQADMDVSELSVADMRSHGPKAKGNPHKPKHLRPTSSAALDKV
EALLTVGGKHWNAYAHAMRMFECKDKVEYLDSTQLHLVAALQIAENRTERAGDGG
VRKS"
216698..217126
/genes="NMA1302"
216698..217126
/genes="NMA1302"
/note="NMA1302, len: 142 aa; unknown, identical to NMA1187
and similar to NMA1866, fasta scores; E(): 1.8e-10, 29.7%
identity in 145 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1302"
/protein_id="CAB84554.1"
/db_xref="GI:7379978"
/db_xref="SPTREMBL:Q9JRW0"
/translation="MGFEKVEHLLPDTVLVDIVIGLAATEOLVKAIGAREKFGKKG
VDRLEALVEAIGEVKTHQLQVGGRELYVPRCGKALQLNRHFRFQEFVKLRDID
KKSGLMAMTKLCPKYGISRTGYTTINEMSRPAAQALF"
complement(217251..217405)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 283.48"
/label=Correia
217423..217427
217434..217979
/genes="NMA1303"
217434..217979
/genes="NMA1303"
/note="NMA1303, possible amidase, len: 181 aa; shows very
weak similarity to TR:P89923 (EMBL:Y11476), ply12,
Bacteriophage 12826 N-acetylmuramoyl-L-alanine amidase
endolysin (bc 3.5.1.28) (257 aa), fasta scores; E(): 0.16,
26.0% identity in 177 aa overlap and to bacterial amidases
(putative) e.g. TR:Q32421 (EMBL:D76414), lytH,
Staphylococcus aureus putative N-acetylmuramoyl-L-alanine
amidase (291 aa), fasta scores; E(): 0.0093, 25.8%
identity in 182 aa overlap. identical to NMA1188 and
highly similar to NMA1864, fasta scores; E(): 0, 96.7%
identity in 181 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="CAB84555.1"
/db_xref="GI:7379979"
/db_xref="SPTREMBL:Q9JUR25"
/translation="MGKTVTLTAGHSNTDPGAVNGSDREADLAQDMRNIVAIRLRDDY
GITVTDGTGKGNMPLREAVKLIRGSDVAIEFTNAAVSKAATGEALSTVKNKRWQY
VLSKAVAKTGWKRGEDGFKPDNAGQHSRLAYAQAGGIVPEPFIISNDTDLALPKTT
KWICRATADATAMELGAARV"
218185..218349
/genes="NMA1304"
218185..218349
```

```
gene
/genes="NMA1304"
/note="NMA1304, len: 54 aa; unknown, identical to NMA1189
and NMA1863"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1304"
/protein_id="CAB84556.1"
/db_xref="GI:7379980"
/db_xref="SPTREMBL:Q9JRV3"
/translation="MQQVAFYEKMIGLSAKSREASEQADLAFAEAGELANYQEML
KRLQTKSV"
218352..218588
/genes="NMA1305"
218352..218588
/genes="NMA1305"
/note="NMA1305, len: 78 aa; unknown, identical to NMA1190
and NMA1862"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1305"
/protein_id="CAB84557.1"
/db_xref="GI:7379981"
/db_xref="SPTREMBL:Q9JQP9"
/translation="MRILDIFKNPATGVNHSKLMANVACAAGTVKFMVLPDPSABIW
AVLGIQGYAVARSLVSVKQEVENESRETAGE"
218560..218979
/genes="NMA1306"
218560..218979
/genes="NMA1306"
/note="NMA1306, len: 139 aa; unknown, identical to NMA1861
and highly similar to NMA1191, fasta scores; E(): 0, 97.1%
identity in 139 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1306"
/protein_id="CAB84558.1"
/db_xref="GI:7379982"
/db_xref="SPTREMBL:Q9JS11"
/translation="MNLVKLLANNRPQIALIALVGTLLAVSHHOGYKSAFAKQAAVID
KNEKDKAQLLLSAQNYARELEQARAEAKYEVKAHVGMALAKKQAEVSRKTKENK
EENLVITQDRKNAGGCDIGFGHGLQLYKRALGYGN"
218840..218849
/genes="NMA1306"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
218999..219193
/genes="NMA1307"
218999..219193
/genes="NMA1307"
/note="NMA1307, len: 64 aa; unknown, highly similar to
NMA1192, fasta scores; E(): 1.9e-11, 82.8% identity in 29
aa overlap and NMA1860, fasta scores; E(): 1.9e-11, 82.8%
identity in 29 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1307"
/protein_id="CAB84559.1"
/db_xref="GI:7379983"
/db_xref="SPTREMBL:Q9JUH8"
/translation="MPTPPAALMVAPVRPNPPKDGKTVTLLEHAAEFGGYVAELENQ
QAWRDVNSQAAVDGSEGAR"
219178..219182
/genes="NMA1307"
219190..219402
/genes="NMA1308"
219190..219402
/genes="NMA1308"
/note="NMA1308, len: 70 aa; unknown, highly similar to
NMA1195, fasta scores; E(): 3.3e-24, 84.4% identity in 64
aa overlap and NMA1157, fasta scores; E(): 1.7e-11, 47.5%
identity in 61 aa overlap"
/codon_start=1
/transl_table=11
```

```
misc_feature
/label=DUS
complement(216163..216172)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
216223..216717
/genes="NMA1301"
216223..216717
/genes="NMA1301"
/note="NMA1301, len: 164 aa; unknown, similar to part of
TR:Q38494 (EMBL:M64097) bacteriophage Mu protein E16 (195
aa), fasta scores; E(): 2.7e-05, 46.7% identity in 60 aa
overlap. Identical to NMA1186 and similar to NMA1867,
fasta scores; E(): 1.5e-25, 55.5% identity in 137 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1301"
/protein_id="CAB84553.1"
/db_xref="GI:7379977"
/db_xref="SPTREMBL:Q9JS31"
/translation="MYFIGICVNSIYCVYEFEMRRALIAKIAQKELGDDGTyr
AVLERVTDKRSQADMDVSELSVADMRSHGPKAKGNPHKPKHLRPTSSAALDKV
EALLTVGGKHWNAYAHAMRMFECKDKVEYLDSTQLHLVAALQIAENRTERAGDGG
VRKS"
216698..217126
/genes="NMA1302"
216698..217126
/genes="NMA1302"
/note="NMA1302, len: 142 aa; unknown, identical to NMA1187
and similar to NMA1866, fasta scores; E(): 1.8e-10, 29.7%
identity in 145 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1302"
/protein_id="CAB84554.1"
/db_xref="GI:7379978"
/db_xref="SPTREMBL:Q9JRW0"
/translation="MGFEKVEHLLPDTVLVDIVIGLAATEOLVKAIGAREKFGKKG
VDRLEALVEAIGEVKTHQLQVGGRELYVPRCGKALQLNRHFRFQEFVKLRDID
KKSGLMAMTKLCPKYGISRTGYTTINEMSRPAAQALF"
complement(217251..217405)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 283.48"
/label=Correia
217423..217427
217434..217979
/genes="NMA1303"
217434..217979
/genes="NMA1303"
/note="NMA1303, possible amidase, len: 181 aa; shows very
weak similarity to TR:P89923 (EMBL:Y11476), ply12,
Bacteriophage 12826 N-acetylmuramoyl-L-alanine amidase
endolysin (bc 3.5.1.28) (257 aa), fasta scores; E(): 0.16,
26.0% identity in 177 aa overlap and to bacterial amidases
(putative) e.g. TR:Q32421 (EMBL:D76414), lytH,
Staphylococcus aureus putative N-acetylmuramoyl-L-alanine
amidase (291 aa), fasta scores; E(): 0.0093, 25.8%
identity in 182 aa overlap. identical to NMA1188 and
highly similar to NMA1864, fasta scores; E(): 0, 96.7%
identity in 181 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="CAB84555.1"
/db_xref="GI:7379979"
/db_xref="SPTREMBL:Q9JUR25"
/translation="MGKTVTLTAGHSNTDPGAVNGSDREADLAQDMRNIVAIRLRDDY
GITVTDGTGKGNMPLREAVKLIRGSDVAIEFTNAAVSKAATGEALSTVKNKRWQY
VLSKAVAKTGWKRGEDGFKPDNAGQHSRLAYAQAGGIVPEPFIISNDTDLALPKTT
KWICRATADATAMELGAARV"
218185..218349
/genes="NMA1304"
218185..218349
```


| | | | | | | | | | | | |
|-------------------|---|-----|--|------|--|--------------|---|--------------|--|--------------|--|
| CDS | complement(213215..213823) /gene="NMA1293" /note="NMA1293, possible phage anti-repressor protein, len: 202 aa; similar to several e.g. TR:Q38585 (EMBL:X95646), ant, Streptococcus thermophilus bacteriophage Sfi21 anti-repressor protein (287 aa), fasta scores; E(): 6.5e-07, 31.3% identity in 134 aa overlap and TR:AAD25485 (EMBL:AF125520), L0142, bacteriophage 933W hypothetical protein (209 aa), fasta scores; E(): 2.7e-17, 37.8% identity in 201 aa overlap" /codon_start=1 /transl_table=11 /product="putative phage anti-repressor protein" /protein_id="CAB84545.1" /db_xref="GI:7379969" /db_xref="SPTREMBL:Q9JUI6" | RBS | /translation="MAEHRAPFANLSASHFTTTILKIDCLCLMKKR" complement(214548..214551) /gene="NMA1296" complement(214814..214981) /gene="NMA1297" complement(214814..214981) /gene="NMA1297" /note="NMA1297, len: 55 aa; unknown, similar to NMA1295, fasta scores; E(): 1.9e-10, 53.8% identity in 52 aa overlap and NMA1294, fasta scores; E(): 1.1e-08, 43.9% identity in 57 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1297" /protein_id="CAB84549.1" /db_xref="GI:7379973" /db_xref="SPTREMBL:Q9JUI2" | gene | /translation="WVDEKLAENRKRYEQKRVIKKVSFNAETEKELLEYAQNIDFSOW VKSIIKEKIRK" complement(214989..214993) 215205..215387 /gene="NMA1298" 215205..215387 /gene="NMA1298" /note="NMA1298, len: 60 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1298" /protein_id="CAB84550.1" /db_xref="GI:7379974" /db_xref="SPTREMBL:Q9JUI1" | CDS | /translation="MNYKRGHVKSQPPFQQNQKRMPSSENSNAVSDGIFCCRAMPCH LCGVAVIKIWFVSUYF" 215267..215276 /gene="NMA1298" /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(215294..215303) /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS 215493..215496 215504..215695 /gene="NMA1299" 215504..215695 /gene="NMA1299" /note="NMA1299, len: 63 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1299" /protein_id="CAB84551.1" /db_xref="GI:7379975" /db_xref="SPTREMBL:Q9JUI0" /translation="MNTVKEYTAVRERLLNAADYLEVRKDRKTGNIASVEFVPPKIG ARGYGRKRVYKTLVAVDL" 215692..216108 /gene="NMA1300" 215692..216108 /gene="NMA1300" /note="NMA1300, possible membrane protein, len: 138 aa; unknown, contains hydrophobic, possible membrane-spanning regions" /codon_start=1 /transl_table=11 /product="putative membrane protein" /protein_id="CAB84552.1" /db_xref="GI:7379976" /db_xref="SPTREMBL:Q9JUI9" /translation="MSGSEKSNLPADPEQQKMLQEFLELQKDIATVAKOKELESRKDEIQ SHERIALATIEAQKQGEQHQEIFFKVRRLNAVIAISFLIAAVALITALWRKDSVA IEIVKIGGAVALGYPAGINKRQAQTKLEKQRREKDS" 215912..215921 /gene="NMA1300" /note="Core DNA uptake sequence: gccgtctgaa" | misc_feature | | misc_feature | |
| stem_loop gene | complement(213903..213937) /gene="NMA1293" complement(213955..214155) /gene="NMA1294" complement(213955..214155) /gene="NMA1294" /note="NMA1294, len: 66 aa; unknown, similar to NMA1297, fasta scores; E(): 1.4e-08, 43.9% identity in 57 aa overlap and NMA1295, fasta scores; E(): 7.1e-07, 51.1% identity in 47 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1294" /protein_id="CAB84546.1" /db_xref="GI:7379970" /db_xref="SPTREMBL:Q9JUI5" /translation="MDSIKKAEKQYALKRVKTVSFNSEKELDLIKFSENLDPSNVVK SKTRGMESEKPKKAKKNIDK" complement(214159..214164) 214236..214245 /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(214264..214273) /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(214356..214541) /gene="NMA1295" complement(214356..214541) /gene="NMA1295" /note="NMA1295, len: 61 aa; unknown, similar to NMA1297, fasta scores; E(): 2.7e-09, 53.8% identity in 52 aa overlap and NMA1294, fasta scores; E(): 4.1e-06, 51.1% identity in 47 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1295" /protein_id="CAB84547.1" /db_xref="GI:7379971" /db_xref="SPTREMBL:Q9JUI4" /translation="MPDEKTLKAKQAEYQRYDQKRLMKTVSFHLERELVDFMQKE IPDFSNNVKSVMKEMK" complement(214520..214621) /gene="NMA1296" complement(214520..214621) /gene="NMA1296" /note="NMA1296, len: 33 aa; unknown" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1296" /protein_id="CAB84548.1" /db_xref="GI:7379972" /db_xref="SPTREMBL:Q9JUI3" | RBS | | gene | | misc_feature | | | | | |


```

RBS      NYAVTMGRKRIPLPLRLQLKQOETQ"
          complement(2050000..2050003)
          /gene="NMA1279"
misc_feature complement(205509..205532)
          /gene="NMA1279"
stem_loop /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
          205727..205762
          /note="Stem loop containing DNA uptake sequences: acgat
          gccgtctgaa aacggc ttcacagcgc atcgt"
          205732..205741
misc_feature /note="Core DNA uptake sequence: gccgtctgaa"
          /label=DUS
          complement(205748..205757)
          /note="Core DNA uptake sequence: gccgtctgaa"
          /label=DUS
          complement(205766..206314)
          /gene="NMA1280"
          complement(205766..206314)
          /gene="NMA1280"
          /note="NMA1280, pseudogene, ABC-transporter ATP-binding
          protein, len: 549 bp; similar to the C-terminal half of
          many putative ABC-transporter ATP-binding proteins e.g.
          TR-054061 (EMBL:AJ225561), exsE, Sinorhizobium meliloti
          putative ABC-transporter ATP-binding protein (606 aa),
          fasta scores; E(): 6.4e-14, 34.9% identity in 172 aa
          overlap and SW:YB67_HAEIN (EMBL:U32825), H11467,
          Haemophilus influenzae putative ABC-transporter
          ATP-binding protein (589 aa), fasta scores; E(): 0, 55.3%
          identity in 179 aa overlap. Note that the N-terminal half
          of this protein appears to be NMA1331, beyond the
          intervening prophage. Contains Pfam match to entry PF00005
          ABC_tran, ABC transporter, PS00211 ABC transporters family
          signature and PS00017 ATP/GTP-binding site motif A
          (P-loop)"
          /codon_start=1
          /pseudo
          /transl_table=11
          /product="ABC-transporter ATP-binding protein
          (pseudogene)"
          complement(205781..206284)
          /gene="NMA1280"
misc_feature /note="Pfam match to entry PF00005 ABC_tran, ABC
          transporter, score 104.10, E-value 2.8e-27"
          complement(205952..205996)
          /gene="NMA1280"
misc_feature /note="PS00211 ABC transporters family signature"
          complement(206240..206263)
          /gene="NMA1280"
          /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
          206315..235513
          /note="possible prophage pm2; boundaries defined by
          NMA1280 and NMA1331, which appear to be the two halves of
          the gene containing the integration site"
          /label=pm2
          206543..206548
          /gene="NMA1280"
          206557..206742
          /gene="NMA1281"
          206557..206742
          /gene="NMA1281"
          /note="NMA1281, len: 61 aa; unknown"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein NMA1281"
          /protein_id="CAB84533.1"
          /db_xref="GI:7379957"
          /db_xref="SPTREMBL:Q9JUJ8"
          /translation="MDLINTPAINTAYDGDVVLGITQIYKRCVTVIFRRYKQFAR
          YTLHTFDKEVIAYTQDI"
          complement(206938..207819)
          /gene="NMA1282"
          complement(206938..207819)
          /gene="NMA1282"
          /note="NMA1282, possible DNA-binding protein, len: 293 aa;

```

```

unknown, contains probable helix-turn-helix at aa 42-63
(score 1152, +3.11 SD)"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="CAB84534.1"
/db_xref="GI:7379958"
/db_xref="SPTREMBL:Q9JUJ7"
/translation="MPALTVTWMIYTKLVIFCVGMERIDIALNLNVEERSRLGYSRK
NFAEQTCSAESRLRYLSEQGVNI PADFLAAVLQGVLDIOQVFTGHSANLANVASDDL
KKLDDAIGKSNVNOINGNSNAVTSAGSVVHOINTONHVIRTRVDSKPKCKEHTLE
QASKLOQLVKQVAAAEETAKRPSKIRAIWASLNHACKVPSYKLTALSDYDKAEETLR
KWLGLSNTATSKNNDPWRKKKIYIILNKVQLELDWLKSYLEKNSVESLTELSD
EDLQRTYAAVSTKKRKK"
207858..207862
207867..208118
/gene="NMA1283"
207867..208118
/gene="NMA1283"
/note="NMA1283, len: 83 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1283"
/protein_id="CAB84535.1"
/db_xref="GI:7379959"
/db_xref="SPTREMBL:Q9JUJ6"
/translation="MQANEIYKLEEKNISARMLASALGVTNQSVSEVIRNGRSGKRI
AEATIAKVIQDLVEVFFHYKTKTDCRKKIAELKEMLVG"
208107..208110
/gene="NMA1283"
208120..210093
/gene="NMA1284"
208120..210093
/gene="NMA1284"
/note="NMA1284,
similar to TR-Q38013 (EMBL:X87627) Bacteriophage D3112
transposase A (690 aa), fasta scores; E(): 0, 34.0%
identity in 623 aa overlap. Similar to NMA1882, fasta
scores; E(): 0, 41.3% identity in 656 aa overlap. Contains
Pfam match to entry PF00552 integrase, Integrase and
PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative phage transposase"
/protein_id="CAB84536.1"
/db_xref="GI:7379960"
/db_xref="SPTREMBL:Q9JUJ5"
/translation="MEYNDLKSLLNFCGLPNWPTVIGLRKKAQRDGMQTRRRCKGGG
VEALPIEIRAAIMKRSQDELAEMKPTLPQVRPGTAMSAQALAEAAKLNEKORSVA
DARCVAVVAVLGIKYEYDCSAKAAVAQFLGLLAEKLDAYTLGNLEKANDRSKAYG
ERTLDGWI SAYLKAENATERLVALAPKTKAVKPIESYGLPMFQPHNIPSPAPKLAH
SYRRFQWAEANNPVDNPNLSWRRWDKLPIMOGRKTKGAAVKSLPLVYKRDV
GALKPDVWITGDGSHFKAQVAHPVHGPRPKEVTIIDGCTRFVVGFSVLSACVAV
SDALRIGVKHFGILPIIYSDNGGGQTKTDHETITGITSRLGIRHERGIGACNFGGRI
IERSWKDNLIEAMRQVETFPASGMDSTKMLMKKMSAFNALEKGLDLEEQQYLLK
KLFWSRFLIADVVYKIDENNNRPHGELPRHPDGGHSPKAYREMRLEQDGIAPDMLSA
EELATFMFQVNRKVQRGWLDLEFNNSYFSTELAEYHKDEVRSYDLSASAVNVFMD
GEFTITKAQDGNTRFAPPTARIDQLAEKRRKGIKRAENAIKLANAEVNPALQEAQVW
DELGLHGGNVIEAEYAVLPKTDGDFVLEADR"
209185..209208
/gene="NMA1284"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
209236..209304
/gene="NMA1284"
/note="NMA1284"
/note="Pfam match to entry PF00552 integrase, Integrase,
score 17.10, E-value 0.00078"
210094..210097
210102..210278
/gene="NMA1285"
210102..210278
/gene="NMA1285"
/note="NMA1285, len: 58 aa; unknown"
/codon_start=1

```

```

misc_feature
misc_feature
RBS
gene
CDS

```

/protein_id="CAB84527.1"
/db_xref="GI:7379952"
/db_xref="SPTREMBL:Q9JUK3"
/translation="MIKISTREDAGSVVVKDLTDPNSIRLALRPDNASDFQAQHFYFRL
QAAAYQNCIMHPENAAEAAPKGEYQACASDTRNRWFRVPTSYENGVLTVNHTPLS
NSYYAYFPFYEEQHNLNLGDAQSGGLCRIDDLGSTVQGRDINLLTIGNQVESDLKI
WITARQHPGETMAEFWEGLLGLLDQDPDPTARTLLDRATFYIVNNPDGSGALNGLR
TAAAGANLANREWENPTLEKSPVFFVREKMLETGVFLDIDHGDEGLPTIFVAGTEGV
PYNPRISALETQFTALLNASPDFODEYGYEKDAPGQANMTLATNWGVNRENCLAYT
LEMPKDNANLPDDDFGHWGQRSRLRGLGEAALSAILNIIGDLR"
complement(201190..201435)
/gene="NMA1274"
/notes="Pfam match to entry PF00246 Zn_carboxipept, Zinc
carboxypeptidase, score 24.70, E-value 1.1e-05"
complement(201830..201833)
201915..201950
/notes="Stem loop containing DNA uptake sequences: cggat
gcgctctgaa gattgt ttcagacggc gtttg"
201920..201929
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(201936..201945)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
201983..201986
201992..202888
/gene="argB"
201992..202888
/gene="argB"
/EC_number="2.7.2.8"
/notes="NMA1275, argB, probable acetylglutamate kinase,
len: 298 aa; similar to many e.g. SW:ARGB_CORGL
(EMBL:X86157), argB, Corynebacterium glutamicum
acetylglutamate kinase (EC 2.7.2.8) (294 aa), fasta
scores; E(): 0, 45.7% identity in 276 aa overlap"
/codon_start=1
/transl_table=11
/product="putative acetylglutamate kinase"
/protein_id="CAB84528.1"
/db_xref="GI:7379953"
/db_xref="SPTREMBL:Q9JUK2"
/translation="MESENIIISAADKARILAEALPYIRFSGSVAVIKYGGNAMTEPA
LKEGFARDVLLKLVGIHPVIVHGGGPQINAMLEKYKKGFEVQGMRTVDKAMDIVE
MYLGHVNRKEIVSMINTYGGHVGVSGRDDHFITAKKLLIDTPEONGVDIGOVTVES
ITGLVKGLIERGCIPIVPAPVGVGEGKEAFNADLVACKLAEEVNAEKLLAMNTIAG
VMDKGNLLTKLTPKRIDELADIAGTLYGGLPKIASAVEAANGVKATHIIDRPVNA
LLEIFIDAGISMLGGGEDA"
complement(202987..203141)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 271.89"
/label=Correia
203171..203190
/notes">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=dRS3
complement(203191..203251)
/notes="85% identity to consensus
CTAGGACGTAAATCTAAAGAACCGTTTACCCGATAGTTCCCGACCACGACAGAC
CTAG"
complement(203242..203396)
/notes="Correia element; hmms hit to HMM Correia (1 -
156), score: 278.56"
/label=Correia
complement(203397..203404)
/notes="partial RS15"
203405..203424
/notes">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=dRS3
complement(203522..204028)
/gene="NMA1276"
complement(203522..204028)
/gene="NMA1276"
/notes="NMA1276, possible membrane protein, len: 168 aa;
unknown, contains hydrophobic, possible membrane-spanning

regions. Lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84529.1"
/db_xref="GI:7379954"
/db_xref="SPTREMBL:Q9JUK1"
/translation="MLIEQSVANHKVSVSHSAFTMKWLLNIMMRPTKFSVMVNTLLFI
VICSSPFDLLVOLCTILFHSOKIYFTLLFLFNFVTKSIYMAIIVPILYFTTKKY
YFYSRKVILLLSIALSIYFSFMDYFFFSYSDNLSETETPLHLXIPIIINFFSLVSN
FILSFINK"
204132..204165
/notes="88% identity to consensus
GACGGCGCATAGTTCCTCGCGGACAGACCTAG"
/label=RS64
204166..204185
/notes">= 90% match to ATTCCCNNNNNNGGGAAT"
/label=dRS3
204281..204323
/notes="Stem loop containing DNA uptake sequences: tcaat
gcgctctgaa cgcgaaatcggc ttcagacggc atgcg"
204286..204295
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(204309..204318)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(204325..204993)
/gene="NMA1278"
complement(204325..204993)
/gene="NMA1278"
/notes="NMA1278, len: 222 aa; unknown, similar to bacterial
hypothetical proteins e.g. TR:O33611 (EMBL:AB004855)
Streptomyces cyaneus protein possibly involved in
morphological differentiation (277 aa), fasta scores; E():
1.5e-15, 29.1% identity in 227 aa overlap and TR:O69629
(EMBL:AL02121) Mycobacterium tuberculosis hypothetical
protein (287 aa), fasta scores; E(): 2.1e-15, 30.2%
identity in 222 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1278"
/protein_id="CAB84530.1"
/db_xref="GI:7379955"
/db_xref="SPTREMBL:Q9JUK0"
/translation="MKNLAIFDLNLTINTSDHAWPOYLKIKGLDYAAETEQAQNEKF
YRDYQNCIDIDAFKLHPLARLYSEELAEHREFMAEYIIPHISPMQMLVQSHQ
MAGDETLVISTNEFTITPVCRLFGITNIGTQESGDRGTGNTGTPSPKEKIT
RLNQWLAERGETLESYGTIFYSDSKNDPLLLRLVDEPVAVNPDAELEKEKKGWVP
LNFK"
complement(204990..205721)
/gene="NMA1279"
complement(204990..205721)
/gene="NMA1279"
/notes="NMA1279, len: 243 aa; similar to SW:YFGE_ECOLI
(EMBL:AE000336), yfge, Escherichia coli hypothetical
protein (248 aa), fasta scores; E(): 3.2e-19, 31.3%
identity in 230 aa overlap and SW:YFGE_HAEIN
(EMBL:U32802), HIL225.1, Haemophilus influenzae
hypothetical protein (231 aa), fasta scores; E(): 5.8e-17,
30.1% identity in 226 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop), which may be
fortuitous"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1279"
/protein_id="CAB84531.1"
/db_xref="GI:7379956"
/db_xref="SPTREMBL:Q9JUK9"
/translation="MIKHAKTINKPVYHPSECRVTNQLIFDFAAHDYPSDFKFLGTE
NAELVYVLRHKHQFIYVWGEEGASHLLQAWAQAALGAANAYIDAASMPFLTDAA
FAEYLAVDQVEKLGNEQALLSIFNFRNRSKGFLLIGSEYTPQOLVIREDLRTM
AYCLYEVKPLTDQEKALLASMAAAARQVTDSEIFEYLLKHWRRMDSLMMMLDITLD

```

/misc_feature
/note="pfam match to entry PF00696 aakinase, Aspartate
kinases, Glutamate kinases and Gamma glutamate
phospho-reductases, score 71.30, E-value 2.1e-17"
196505..196514
/codon_start=1
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(196795..196798)
RBS
197259..197264
gene
197272..199825
/gene="leuA"
CDS
197272..199825
/gene="leuA"
/EC_number="4.1.3.12"
/note="NMA1270, leuA, probable 2-isopropylmalate synthase,
len: 597 aa; similar to many e.g. SW:LEU1_ECOLI
(EMBL:D10483), leuA, Escherichia coli 2-isopropylmalate
synthase (EC 4.1.3.12) (522 aa), fasta scores; E(): 0,
46.9% identity in 501 aa overlap. Contains pfam match to
entry PF00682 HMGL-like, HMGL-like, PS00815
Alpha-isopropylmalate and homocitrate synthases signature
1 and PS00816 Alpha-isopropylmalate and homocitrate
synthases signature 2"
/codon_start=1
/transl_table=11
/product="putative 2-isopropylmalate synthase"
/protein_id="CAB84524.1"
/db_xref="GI:7379949"
/db_xref="SPTREMBL:Q9JUK6"
/translation="MTQANRVIIIFDTTLRDSQSPGAAMTKEEKIRVARQLEKLGVDI
TEAGFAASPGDFEAVNAIAKTIITKSTVCSLSRAIERDIOAGAEVAPAPKPKRIHTFI
ATSDIHMEYKLMKPKQVIEAAVAKIAREYTDVDFSCEDALRSETDFLAEICGAV
IEAGATTINIPDTGVGISPYKTEFFERELIVKTPNGKVVNSACHNDLGLAVANSLA
ALKGAARQVETVNGLGERAGNASVEEIVALKVRHDLFGLDETQIDITQIVPSSKLVS
TITGYPQPNKAIVGANAFSHESGIHQDVLKHKRETYEIMSAGESVGWATNRLSLKLS
GRNAFKTLADLGELESEEAALNAAFARFKELADKREIFDEDLHALVSDMSGWNAE
SYKFSIKSTETGEEPRADIVFSIKGEKRSATGSGPVDIAIFKATESVAOSGAALQ
IYSVNAVTVQTSQGETSVRLARGNRVVGOGADTVLVATAKAYLSALSLEFGSAK
PKAOGSGTI"
197308..198144
/gene="leuA"
/misc_feature
/note="pfam match to entry PF00682 HMGL-like, HMGL-like,
score 410.80, E-value 1.3e-119"
197311..197361
/gene="leuA"
/misc_feature
/note="PS00815 Alpha-isopropylmalate and homocitrate
synthases signature 1"
197872..197913
/gene="leuA"
/misc_feature
/note="PS00816 Alpha-isopropylmalate and homocitrate
synthases signature 2"
198838..198879
/note="stem loop containing DNA uptake sequences: acgat
gccgtctgaa gcataaaaaggc ttcacagcgc attgc"
198843..198852
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(198865..198874)
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(198875..198877)
/label=DUS
complement(198885..198927)
/note="Correia element; hmufs hit to HMM Correia (114 -
156), score: 51.92"
/label=Correia
complement(198929..198991)
/note="Correia element; hmufs hit to HMM Correia (1 - 62),
score: 78.71"
/label=Correia
RBS
199016..199019
gene
199025..199687
/gene="NMA1272"
CDS
199025..199687
/gene="NMA1272"

```

```

/note="NMA1272, len: 220 aa; unknown, shows weak
similarity to SW:YECA_ECOLI (EMBL:AE000284), Yeca,
Escherichia coli hypothetical protein (221 aa), fasta
scores; E(): 5.5e-12, 27.2% identity in 228 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1272"
/protein_id="CAB84525.1"
/db_xref="GI:7379950"
/db_xref="SPTREMBL:Q9JUK5"
/translation="MDSRKFTASRRSLSELIDAKSEQNTMRCDEVOGFWTALLSGP
DKLTPDLWLPVLGDSQFTAAERSEIERLVLAMAMETTAAMSDKKLPDLWLYENEDG
GSDFTWCNAYLYGLDIPVTDWFAVDDEAEELFPIALGGIYDEEDNGTIRLOFT
EGEELAELESELPALADIYRWQAVINPQIVVRREGEKTRNDPCCGSGRRIKACCG
KN"
199720..199729
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(199770..200621)
/gene="lgt"
CDS
complement(199770..200621)
/gene="lgt"
/EC_number="2.4.99.-"
/note="NMA1273, lgt, probable prolipoprotein
diacylglyceryl transferase, len: 283 aa; similar to many
e.g. SW:LGT_ECOLI (EMBL:012289), lgt, Escherichia coli
prolipoprotein diacylglyceryl transferase (EC 2.4.99.-)
(291 aa), fasta scores; E(): 0, 51.8% identity in 272 aa
overlap. Contains PS01311 Prolipoprotein diacylglyceryl
transferase signature. Contains hydrophobic, probable
membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative prolipoprotein diacylglyceryl
transferase"
/protein_id="CAB84526.1"
/db_xref="GI:7379951"
/db_xref="SPTREMBL:Q9JUK4"
/translation="MITHPOQDPVLISIGSLAVRWYALSYILGFTLFTFLGRRRIAQG
LSVFTKESLDLDFLWGLIGLVILGGRLGYLFKFSYDLAHLPLDFKWEQMSFHGGF
LGVVIAIWLFGKTHGIGIKLMDTVAPLPLGLASGRIGNGLMGRVTDINAFWA
MGFFQARVEDLEAAAHNPQWMLPRHPQSOLYQFALGICLFAVVVLPFSKKQ
RPTQOASLFLGGYGIQFRFIAEFARQDDYLGILLTLGLSMQWLVSVPMLVIGIVGVR
FGMKKQH"
complement(200172..200210)
/gene="lgt"
/note="PS01311 Prolipoprotein diacylglyceryl transferase
signature"
200638..200678
/misc_feature
/note="stem loop containing DNA uptake sequences: caaat
gccgtctgaa aggcgatgagc ttcacagcgc atttc"
200643..200652
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(200664..200673)
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(200692..201822)
/gene="NMA1274"
complement(200692..201822)
/gene="NMA1274"
/note="NMA1274, len: 376 aa; unknown, similar to
SW:YFG2_ZYMMO (EMBL:I09649), zm2, Zymomonas mobilis
hypothetical protein (380 aa), fasta scores; E(): 0, 47.1%
identity in 374 aa overlap. Also similar to parts of
Caenorhabditis elegans hypothetical proteins e.g.
TR-076373 (EMBL:AF067618), F56H1.5, Caenorhabditis elegans
hypothetical protein (1192 aa), fasta scores; E(): 5e-11,
27.1% identity in 317 aa overlap. Contains pfam match to
entry PF00246 Zn_carboxypeptidase"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1274"

```

aa; similar to many e.g. SW:FTSK_ECOLI (EMBL:249932),
ftsK, Escherichia coli cell division protein (1329 aa),
fasta scores; E(): 0, 40.0% identity in 1015 aa overlap,
the E.coli protein being approx. 300 aa longer at the
N-terminus. Note that this means most of the predicted
membrane spans in the E.coli homologue are absent from
this protein. Similar over the C-terminal approx. 500 aa
to NMA1527, fasta scores; E(): 0, 57.9% identity in 556 aa
overlap. Contains Pfam match to entry PF01580
FtsK_SpoIIIE, FtsK/SpoIIIE family and PS00017
ATP/GTP-binding site motif A (P-loop). Contains a probable
N-terminal signal sequence and a hydrophobic, probable
membrane-spanning region"
/codon_start=1
/transl_table=11
/product="putative cell-division protein"
/protein_id="CAB84521.1"
/db_xref="GI:7379946"
/db_xref="SPTREMBL:Q9JUK9"
/translation="MEWLVILVILLALAGLFFVRAQSEREMREVSANQKKGEKQA
ELPEIRKGMDFPELALMLFHAVKTAVYWLFGVRFRCNRYLAHESEDPRPVPPASAN
RADVPASDGSYSDSGNGTEAEAEAEAEADTEDIATAVIDNRRIPFDRSIAEG
LMPSEESI SVPRVFKEITLEEATRALNSAALRETKRYIDAFENETAVPKYRVSDT
PMGELQIGLDDVLQRTYSRMDADKEAFSESADYGFEPYFEKQHPSPAFSAVKAENA
RNAFRRHAGDGGKQAEAKSPDYSOGSVSDGTAVRDARRVSNVNLKEPNKATVSAEA
RISRLIPESRTVVGKRDVEMPSSETENVETVTSVCGYGGPVDETADHIEEPAAPDA
WVVEPEVPKVPAPADIPPPPVSEIYNTYEPFAGFEQVQSRIAETDHLADDVLN
GMOEETAANDSGSEGAERSQYLSEAFGHDSQAVCPENPSEPSRRASDT
EADGAFQSETGAVSEHLPTDLLPLFPNGATQTEELSENSITIEKLAEFVKV
KVVYDSGPVITRYETIEPDVGVGNSVLNLEKDLARSLGASIRVVEITLGTGML
ELNPKRQMLRLSEIFNSPEFAESKSLTALGODITGQPVVDLKGAPHLHVGTTG
SGKSVGNMILSMLFKAAPEDVIMIDPKMLELSYEGIPHLAPVVTDMKLAANA
LNCVNEMKRYRLMSFMGVNRLAGNQKIAEAAARGEKIGNFPFSLTPDNPEPLEKLP
FIVVVDFAFLMWTAGKLTIEELIALAOKARAAGIHLILATORPSVDVITGLIKANI
PTRIAFQSGPYDDILQMGAEINLGGQMLFLPPGTAYPQRVHGFAGASDEVRV
VEYLKQFGEPPYVDDILSGMSDDLIGISRSQGETDPMYDEAVSVVLKTRKASISGV
ORALRIGYNAARLIDOMEAEGIVSAPEHNGNRTLILVPLDNA"
191618 .191627
/gene="NMA1266"
/note="Core DNA uptake sequence: gccgtctgaa"
complement(192047. .192056)
/label=DUS
/note="Core DNA uptake sequence: gccgtctgaa"
192194 .192203
/gene="NMA1266"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
192620 .192629
/gene="NMA1266"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
193035 .193667
/gene="NMA1266"
/note="Ffam match to entry PF01580 FtsK_SpoIIIE,
FtsK/SpoIIIE family, score 304.00, E-value 1.8e-87"
193167 .193190
/gene="NMA1266"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
194191 .194226
/note="Stem loop containing DNA uptake sequences: gaaat
gccgtctgaa gactgt ttcagacggc atttt"
194196 .194205
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(194212. .194221)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
194227 .194382
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 278.52"
complement(194402. .195664)
gene

/gene="proA"
complement(194402. .195664)
/gene="proA"
/EC_number="1.2.1.41"
/note="NMA1267, proA, probable gamma-glutamyl phosphate
reductase, len: 420 aa; similar to many e.g. TR:P96489
(EMBL:X92418), proA, Streptococcus thermophilus
gamma-glutamyl phosphate reductase (EC 1.2.1.41) (416 aa),
fasta scores; E(): 0, 50.0% identity in 412 aa overlap.
Contains Pfam match to entry PF00696 aakinas, Aspartate
kinases, Glutamate kinases and Gamma glutamate
phospho-reductases and PS01223 Gamma-glutamyl phosphate
reductase signature"
/codon_start=1
/transl_table=11
/product="putative gamma-glutamyl phosphate reductase"
/protein_id="CAB84522.1"
/db_xref="GI:7379947"
/db_xref="SPTREMBL:Q9JUK8"
/translation="MSNTQQLALAKAAKSVNTADTEKNRALFAMADSLERAAAADI
LANRODLCAAGNIPESMTDRLLGKRICAMADGIRAVAAPDPVGELETSTLPN
GLEIVKRVAMGVIGIIEYSPNVTDAALALKSGSAVVLRSKGAGTASARAIVAAAL
KTGLAQRIDPDALQIIEDTRESSYEMRAKDYLDLLIPRGAGITRAVENAVVPV
IETGTGIVHYIDKADADPKALRIVYNAKTSRPSVCNSMEVLVHEDIAADLPKLER
ILVNRNIEAGLPPVRELDPQAAARHIGGAAGADDFDTEFLDYILAVKTVASVEEAVG
HIEAHGTHSDGIVTENRHAADYFTTHIDSAVYVNASTRFTDGGEGLCCEMGISTQ
KLHARPGMLKELTSYIYVQGTGVQVRE"
/gene="proA"
complement(194624. .194689)
/gene="proA"
194627 .194636
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(194879. .195508)
/gene="proA"
/note="Ffam match to entry PF00696 aakinas, Aspartate
kinases, Glutamate kinases and Gamma glutamate
phospho-reductases, score 170.60, E-value 2.6e-47"
complement(195670. .195673)
complement(195677. .196786)
/gene="proB"
complement(195677. .196786)
/gene="proB"
/EC_number="2.7.2.11"
/note="NMA1268, proB, probable glutamate 5-kinase, len:
369 aa; similar to many e.g. SW:PROB_BACSU (EMBL:D26044),
proB, Bacillus subtilis glutamate 5-kinase (EC 2.7.2.11)
(365 aa), fasta scores; E(): 0, 42.1% identity in 366 aa
overlap. Contains Pfam match to entry PF00696 aakinas,
Aspartate kinases, Glutamate kinases and Gamma glutamate
phospho-reductases, PS00902 Glutamate 5-kinase signature
and PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative glutamate 5-kinase"
/protein_id="CAB84523.1"
/db_xref="GI:7379948"
/db_xref="SPTREMBL:Q9JUK7"
/translation="MKYKRIVFKVIGTSSITHSDGSLSRGKIQTITRQLAALHHAGHEL
VLVSSCAVAGFGALGFKRKPVKIADKQASAAVGGQLMEEYANLSSDGI VSAQIL
SRADFADKRYONAGGALS VLLQRRAPVITINENDTVSVELKIGDNDTLSAQVAMIQ
ADLLVLLTDIGLYTGNPNSPDVARDLKHIEHNHEIEMAGSGSANGTGMLTIK
GATVATESGPVYVICSLLAPDALAADQADQAFVPRAKGURTKQWLAFTYSESRG
GVYVDEGAELHSEQGLSLLMGIAEGIEGHFSRMDTVTVYSKATKPLGKGRVLFGSA
AAEDLLKRAKGVFTHRDDWISITPEIRILLTEF"
complement(195905. .195928)
/gene="proB"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(196106. .196171)
/gene="proB"
/note="PS00902 Glutamate 5-kinase signature"
complement(196160. .196768)
misc_feature

```

CDS
188708..189244
/gene="NMA1263"
/notes="NMA1263, len: 178 aa; unknown, similar to many bacterial hypothetical proteins e.g. SW:YOKG_BACSU (EMBL:D84432), yqkG, Bacillus subtilis hypothetical protein (185 aa), fasta scores; E(): 6.9e-21, 37.9% identity in 177 aa overlap. Contains Pfam match to entry PF00293 mutT, Bacterial mutT protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1263"
/protein_id="CAB84518.1"
/db_xref="GI:7379943"
/db_xref="SPTREMBL:O9JUL2"
/translation="MDLREVLGGTIYEGGFVSLSRDKVRLPNEGQRIIVIRHGA ACVLATDEKVVLYQWRKYRANQATLSPAGKLDVAGEDAACALRELAETPTT SVRLYSYTAVGCNEKMYLFEAEGRVLSGTLANDEDEITETVLMSEKVRQALAND EIKDKGTLIGLQYLMKD"
/188873..189010
/gene="NMA1263"
/notes="Pfam match to entry PF00293 mutT, Bacterial mutT protein, score 22.30, E-value 7.3e-05"
189589..189625
/complement(189668)..189822
/notes="Correia element; hmms hits to HMM Correia (1 - 156), score: 271.05"
/label=Correia
/complement(189900)..190259
/gene="NMA1264"
/complement(189900)..190259
/gene="NMA1264"
/notes="NMA1264, possible integral membrane protein, len: 119 aa; similar bacterial putative membrane proteins e.g. SW:CRCB_ECOLI (EMBL:AE000167), crCB, Escherichia coli protein which may be involved in chromosome folding (127 aa), fasta scores; E(): 4.3e-13, 42.3% identity in 123 aa overlap. Contains hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84519.1"
/db_xref="GI:7379944"
/db_xref="SPTREMBL:O9JUL1"
/translation="MLSNLISGIALGATFARWLLNLAVPASLPATGNLFANMTGA FLIGFATFVNRHPQMKLLITLFTGLSLETVTTLQSNRPASALANIFLHTA GSLIITMGLKIGTAVK"
190276..190285
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/complement(190288)..190968
/gene="NMA1265"
/complement(190288)..190968
/gene="NMA1265"
/notes="NMA1265, probable integral membrane protein, len: 226 aa; unknown, contains hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84520.1"
/db_xref="GI:7379945"
/db_xref="SPTREMBL:O9JUL0"
/translation="MIFOTVMFSDMVLISVIVLILIAASAPSAFSLARYBSALPL CTVIFSAAGCLNASAGGQALQNMVHLLAVNLVAMVGTSAALWLAALLMLPCLLFA DSAGYAPNALVILPLVNLVNRSLMNLNPPNIFIFVNGFLASAGAILLTGLVLTGILDAANAPFSEILMTTALPVLIFILLAAEAFSLGISTAIFFALKPHWIINTFDNRYL KSDRGIRW"
191133..194177
/gene="NMA1266"
191133..194177
/gene="NMA1266"

```

misc_feature

```

complement(180951. .182081)
/genes="glyA"
/notes="Pfam match to entry PF00464 SHMT, Serine
hydroxymethyltransferase, score 847.00, E-value 4.5e-254"
complement(181392. .181442)
/genes="glyA"
/notes="PS00096 Serine hydroxymethyltransferase
pyridoxal-phosphate attachment site"
complement(182111. .182115)
complement(182583. .184358)
/genes="ggt"
complement(182583. .184358)
/genes="ggt"
/EC_number="2.3.2.2"
/notes="NMA1255, ggt, probable
gamma-glutamyltranspeptidase, len: 591 aa; similar to many
e.g. SW:GGT_BACSU (EMBL:U49358), ggt, Bacillus subtilis
gamma-glutamyltranspeptidase precursor (EC 2.3.2.2) (587
aa), fasta scores; E(): 0, 38.2% identity in 573 aa
overlap. Contains pfam match to entry PF01019
G.glu.transpept, Gamma-glutamyltranspeptidase and an
appropriately positioned PS00013 Prokaryotic membrane
lipoprotein lipid attachment site. Contains a probable
N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative gamma-glutamyltranspeptidase"
/protein_id="CAB84510.1"
/db_xref="GI:7379936"
/db_xref="SPTREMBL:Q9JUL8"
/translacion="MAKTYLLTALIMSMTISGCOVIHANOQKVNTSHAVITGADATPT
EHATGLTQKQVIADEMFVASANPLATOAGVDILKOGGSAADAMVAVOTTLSLVEPOS
SGLGGAVLWNTAKTLTFDGETAPMRATPELFDKDCQPLKFEAVVGGRSVG
TPAIPKMETIHORGLVLPWGKLFDPTRLAKQGFESFRLAISVEQVQHLARIPKT
AAFLPNCVPIQAGSLNLEFADFQAQAQAKALHTGYAQNVISVQNAKDNPNG
QLSLQDLDYQVVERPPVCVTYRIYEVCGMGAPSSGGIAVGQILGILNEFSPNQVGYD
REGLLDGAQRLEAFADRDVYLGDPDFVPTROLISKDYLKHSOLLEQSDKALPSV
SAGDFIHEWVSQATLELPSTSHISTVDKAGNVLSMTTSENAGSTLMANGYLLNNEI
TDFSEPIKQKQVARNVPEKRRPSSWAPTIVFKAGRPYMAIGSPGSRIGIYAKT
IVASHDNMDIQDAISAPLNLRFGSYELRTGTTALQWQALNDGLGYKTDVRELNSGV
QAIIIEPSRLVGGADPRREGRVMGD"
complement(182598. .184136)
/genes="ggt"
/notes="Pfam match to entry PF01019 G.glu.transpept,
Gamma-glutamyltranspeptidase, score 527.30, E-value
1.1e-154"
complement(184302. .184334)
/genes="ggt"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(184364. .184368)
complement(184569. .184792)
/genes="NMA1256"
complement(184569. .184792)
/genes="NMA1256"
/notes="NMA1256, probable pseudogene, len: 224 bp; similar
to many hypothetical proteins, some of which are
glutaredoxin-like e.g. PR:O30824 (EMBL:AF017750)
Haemophilus ducreyi hypothetical protein (107 aa), fasta
scores; E(): 1.7e-10, 73.8% identity in 42 aa overlap. No
remains of a glutaredoxin active site are apparent in this
CDS"
/codon_start=1
/pseudo
/transl_table=11
/product="hypothetical protein NMA1256 (pseudogene)"
complement(184788. .184942)
/notes="Correia element; hmms hits to HMM Correia (1 -
156), score: 284.48"
/label=Correia
184927. .185154
/genes="NMA1257"
184927. .185154

```

misc_feature

```

complement(182598. .184136)
/genes="ggt"
/notes="Pfam match to entry PF01019 G.glu.transpept,
Gamma-glutamyltranspeptidase, score 527.30, E-value
1.1e-154"

```

misc_feature

```

complement(184302. .184334)
/genes="ggt"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(184364. .184368)
complement(184569. .184792)
/genes="NMA1256"
complement(184569. .184792)
/genes="NMA1256"

```

RBS

gene

CDS

```

/notes="NMA1256, probable pseudogene, len: 224 bp; similar
to many hypothetical proteins, some of which are
glutaredoxin-like e.g. PR:O30824 (EMBL:AF017750)
Haemophilus ducreyi hypothetical protein (107 aa), fasta
scores; E(): 1.7e-10, 73.8% identity in 42 aa overlap. No
remains of a glutaredoxin active site are apparent in this
CDS"
/codon_start=1
/pseudo
/transl_table=11
/product="hypothetical protein NMA1256 (pseudogene)"
complement(184788. .184942)
/notes="Correia element; hmms hits to HMM Correia (1 -
156), score: 284.48"
/label=Correia
184927. .185154
/genes="NMA1257"
184927. .185154

```

repeat_unit

```

complement(184788. .184942)
/notes="Correia element; hmms hits to HMM Correia (1 -
156), score: 284.48"
/label=Correia
184927. .185154
/genes="NMA1257"
184927. .185154

```

gene

CDS

```

/genes="NMA1257"
/notes="NMA1257, len: 75 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1257"
/protein_id="CAB84512.1"
/db_xref="GI:7379937"
/db_xref="SPTREMBL:Q9JUL7"
/translacion="MLIHVTYGAENLCVKRDIVNVFNRTAVLKENFAVSKTLDKN
IIFVVIINVKIQDFELTVNQERKAICSQNTVI"
185117. .185121
/genes="NMA1257"
185130. .185348
/genes="NMA1258"
185130. .185348
/genes="NMA1258"
/notes="NMA1258, len: 72 aa; similar to hypothetical
proteins TR:Q92JY4 (EMBL:AE001544) Helicobacter pylori J99
hypothetical protein (76 aa), fasta scores; E(): 1.1e-06,
47.0% identity in 66 aa overlap and TR:O25839
(EMBL:AE000629) Helicobacter pylori 26695 hypothetical
protein (76 aa), fasta scores; E(): 1.3e-06, 45.5%
identity in 66 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1258"
/protein_id="CAB84513.1"
/db_xref="GI:7379938"
/db_xref="SPTREMBL:Q9JUR9"
/translacion="MFPEYRDLISKLQKNSRFRARLDFDEHNELDKKTGLVNNPVTSG
AETIDELKAKLKLKLDLYAILQRAAG"
185367. .185401
/notes="Stem loop containing DNA uptake sequences: aaaat
gccgtctgaa aacct ttcagagcgc atttt"
185372. .185381
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(185387. .185396)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(185507. .186481)
/genes="fbp"
complement(185507. .186481)
/genes="fbp"
/EC_number="3.1.3.11"
/notes="NMA1259, fbp, probable fructose-1,6-bisphosphatase,
len: 324 aa; similar to many e.g. SW:F16P_XANFL
(EMBL:X17252), cfxF, Xanthobacter flavus
fructose-1,6-bisphosphatase (EC 3.1.3.11) (364 aa), fasta
scores; E(): 0, 47.8% identity in 322 aa overlap. Contains
pfam match to entry PF00316 FBpase,
Fructose-1,6-bisphosphatase"
/codon_start=1
/transl_table=11
/product="putative fructose-1,6-bisphosphatase"
/protein_id="CAB84514.1"
/db_xref="GI:7379939"
/db_xref="SPTREMBL:Q9JUL6"
/translacion="MDTLRELPHLQONQLPEALGGVLLSVVSACTEINAKVRIGAL
AGVLGMAGTNIQGGDKKLDVINIMIDLKANPAVAGLASEEEDSFVSAGENGRY
LVLPDPLDGSSNIDVNSVGTIFSILKEPEGALATESFLQTGROOLAAGYVLYGPTQY
LVFTFGHGVMTLNAENEFVLTREKPKVPPESTKEFALNMSRRHLPPVOQYIDELL
AGETGRGKNYNMRKVAEMVAEIIHRLMRGGVFMYPODKRPSKPKRLMYEANPMA
LILEQAGASASNAQDILGIRPESLHQRVAVIMGSSSEEDVILNRLHSK"
complement(185537. .186310)
/genes="fbp"
/notes="Pfam match to entry PF00316 FBpase,
Fructose-1,6-bisphosphatase, score 467.30, E-value
1.2e-136"
186508. .186517
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(186710. .186719)

```

misc_feature

misc_feature

misc_feature


```

/ gene="NMA1242"
complement(167583..168601)
/ gene="NMA1242"
/ note="NMA1242, pseudogene, probable transposase, len:
1019 bp; highly similar to many e.g. TR:CA844967
(EMBL:AJ242841) Neisseria meningitidis strain 22491
putative transposase for IS1106A3 (335 aa), fasta scores;
E(): 0, 99.0% identity in 96 aa overlap"
/ codon_start=1
/ pseudo
/ transl_table=11
/ product="putative transposase (pseudogene)"
168154..168308
/ note="Correia element; hmms hit to HMM Correia (1 -
156), score: 284.48"
/ label=Correia
complement(168778..168787)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
168877..168912
/ note="Stem loop containing DNA uptake sequences: gaaat
gccgtctgaa agcggg ttacagcggc atagc"
168882..168891
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(168898..168907)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(168915..169970)
/ gene="NMA1243"
complement(168915..169970)
/ gene="NMA1243"
/ note="NMA1243, probable sulphate-binding protein, len:
351 aa; similar to many e.g. SW:SUBI_ECOLI (EMBL:X02519),
sbp, Escherichia coli sulfate-binding protein precursor
(329 aa), fasta scores; E(): 0, 49.1% identity in 316 aa
overlap. Contains Pfam match to entry PF01100
Sulphate_bind, Prokaryotic sulphate- and
thiosulphate-binding protein and an appropriately
positioned PS00013 Prokaryotic membrane lipoprotein lipid
attachment site. Contains a probable N-terminal signal
sequence"
/ codon_start=1
/ transl_table=11
/ product="putative sulphate-binding protein"
/ protein_id="CA884498.1"
/ db_xref="GI:7379925"
/ db_xref="SPTREMBL:Q9JUM0"
/ translation="MKTYPALYTAALLTACSPAADSNHPSGQNPANTSDGKNITL
LNASYDVARDFYKEYNPLFIKTYQSEHPGTSVSTQSHGSSKQALSVANGLOADVVT
MNQSSDIDLEKKGLVEKGWQALPDHAAPYTTMFLVRKNPKQIRDNHDLAKGV
NIVTANPKTSNGRYAFLGAYGYCLKTNGNEQEAQKLVASILKNTVPFENGGAATT
TFTQRNIGDVLITFENENAVYVKLTQCGFEIVPSTIISAESPVAVVNSVAKKGTQ
KTRAYLEYLWSEPAQEAASLYLRNPEVLARHKADFPDLDTFSPPEKFGGWDNIM
KTFADGGIFDRLTAQR"
complement(168921..169970)
/ gene="NMA1243"
/ note="Pfam match to entry PF01100 Sulphate_bind,
prokaryotic sulphate- and thiosulphate-binding protein,
score 485.20, E-value 5.2e-142"
complement(169920..169952)
/ gene="NMA1243"
/ note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(169978..169981)
/ gene="NMA1244"
complement(170076..170558)
/ gene="NMA1244"
complement(170076..170558)
/ note="NMA1244, len: 160 aa; similar to TR:P72792
(EMBL:D90900) Synechocystis sp. (strain PCC 6803)
hypothetical protein (160 aa), fasta scores; E():
28.8% identity in 160 aa overlap. Shows very weak

/ gene="NMA1242"
complement(167583..168601)
/ gene="NMA1242"
/ note="NMA1242, pseudogene, probable transposase, len:
1019 bp; highly similar to many e.g. TR:CA844967
(EMBL:AJ242841) Neisseria meningitidis strain 22491
putative transposase for IS1106A3 (335 aa), fasta scores;
E(): 0, 99.0% identity in 96 aa overlap"
/ codon_start=1
/ pseudo
/ transl_table=11
/ product="putative transposase (pseudogene)"
168154..168308
/ note="Correia element; hmms hit to HMM Correia (1 -
156), score: 284.48"
/ label=Correia
complement(168778..168787)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
168877..168912
/ note="Stem loop containing DNA uptake sequences: gaaat
gccgtctgaa agcggg ttacagcggc atagc"
168882..168891
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(168898..168907)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(168915..169970)
/ gene="NMA1243"
complement(168915..169970)
/ gene="NMA1243"
/ note="NMA1243, probable sulphate-binding protein, len:
351 aa; similar to many e.g. SW:SUBI_ECOLI (EMBL:X02519),
sbp, Escherichia coli sulfate-binding protein precursor
(329 aa), fasta scores; E(): 0, 49.1% identity in 316 aa
overlap. Contains Pfam match to entry PF01100
Sulphate_bind, Prokaryotic sulphate- and
thiosulphate-binding protein and an appropriately
positioned PS00013 Prokaryotic membrane lipoprotein lipid
attachment site. Contains a probable N-terminal signal
sequence"
/ codon_start=1
/ transl_table=11
/ product="putative sulphate-binding protein"
/ protein_id="CA884498.1"
/ db_xref="GI:7379925"
/ db_xref="SPTREMBL:Q9JUM0"
/ translation="MKTYPALYTAALLTACSPAADSNHPSGQNPANTSDGKNITL
LNASYDVARDFYKEYNPLFIKTYQSEHPGTSVSTQSHGSSKQALSVANGLOADVVT
MNQSSDIDLEKKGLVEKGWQALPDHAAPYTTMFLVRKNPKQIRDNHDLAKGV
NIVTANPKTSNGRYAFLGAYGYCLKTNGNEQEAQKLVASILKNTVPFENGGAATT
TFTQRNIGDVLITFENENAVYVKLTQCGFEIVPSTIISAESPVAVVNSVAKKGTQ
KTRAYLEYLWSEPAQEAASLYLRNPEVLARHKADFPDLDTFSPPEKFGGWDNIM
KTFADGGIFDRLTAQR"
complement(168921..169970)
/ gene="NMA1243"
/ note="Pfam match to entry PF01100 Sulphate_bind,
prokaryotic sulphate- and thiosulphate-binding protein,
score 485.20, E-value 5.2e-142"
complement(169920..169952)
/ gene="NMA1243"
/ note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(169978..169981)
/ gene="NMA1244"
complement(170076..170558)
/ gene="NMA1244"
complement(170076..170558)
/ note="NMA1244, len: 160 aa; similar to TR:P72792
(EMBL:D90900) Synechocystis sp. (strain PCC 6803)
hypothetical protein (160 aa), fasta scores; E():
28.8% identity in 160 aa overlap. Shows very weak

/ gene="NMA1244"
complement(170151..170540)
/ gene="NMA1244"
/ note="Pfam match to entry PF00583 Acetyltransf,
Acetyltransferase (GNAT) family, score 32.20, E-value
1.2e-05"
complement(170576..170585)
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(170605..171741)
/ gene="purK"
complement(170605..171741)
/ gene="purK"
/ EC_number="4.1.1.21"
/ note="NMA1245, purK, probable
phosphoribosylaminoimidazole carboxylase ATPase subunit,
len: 378 aa; previously sequenced from Neisseria
meningitidis 22491 as TR:CA844975 (EMBL:AJ242841). Similar
to many e.g. SW:PURK_BACSU (EMBL:J02732), purK, Bacillus
subtilis phosphoribosylaminoimidazole carboxylase ATPase
subunit (EC 4.1.1.21) (379 aa), fasta scores; E(): 0,
44.5% identity in 375 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="putative phosphoribosylaminoimidazole
carboxylase ATPase subunit"
/ protein_id="CA884500.1"
/ db_xref="GI:7379927"
/ db_xref="SPTREMBL:Q9XAY8"
/ translation="MKNISPPAMLILGGQGLGRMTVAAKTMGYKVTVLDPNPNA
PAAEFADRHLCAPEDNQTALAEELAKCAAVTEFENVADAMFLAKHTNVSQDQVVA
IAQNRIOEKAWIRKAGLQTAPOAICKAEDITEESIOFLPGILKTATLGYDGKQIEV
KTVDLEKAAFAHRGVDCVLEKWDLRGEISVIVCLNNDNVQTFDPAENIHENGILA
YSIVPAELSDADIQQAQRMQARLADENLYGVGLAVEMFVGDTHVELVNETAPRPHNS
GHVTVACADQFCQQVRLCMNPPADTKLLSSCCCMANILGDVWQEDGGEDPDPFPIQS
RPOAHLHLYGKTKAHRKRMGHFTILTSDSDTAQFEAKKLHQSL"
complement(171149..171158)
/ gene="purK"
/ note="Core DNA uptake sequence: gccgtctgaa"
/ label=DUS
complement(171738..172283)
/ gene="NMA1246"
complement(171738..172283)
/ gene="NMA1246"
/ note="NMA1246, len: 181 aa; unknown, previously sequenced
from Neisseria meningitidis 22491 as TR:CA844974
(EMBL:AJ242841). Lies within a region of unusually low GC
content"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein NMA1246"
/ protein_id="CA884501.1"
/ db_xref="GI:7379928"
/ db_xref="SPTREMBL:Q9XAY9"
/ translation="MTYEKLYELRQTLRPTTKDGLYTDNEKSREILTVRWSGNTENPK
NFAVAIGNPSKANDERSDKTLTOLARFLDMYGFNFKMLNIFSSYSTQOQTGIRANT
OTDFSFKCGLCEDADMILLAWGCTDSAYRDEKRNLEFLKAEKFMKVCISETGNS
DTRHPSRISYSYLVQVFEESA"

```

CDS

repeat_unit

misc_feature

stem_loop

misc_feature

misc_feature

gene

CDS

misc_feature

misc_feature

RBS

gene

CDS


```

/misc_feature
/translation="MNTETIETATDKVSNQAMRSVLVMEQNGETYSERWVERGR
QAVRTMEGFEELRALIILKEHTGRFMEIVKVSQFGLGIAETSRILMSATRRFATPO
MOKAAPKLDLCKSKLLLEEDVTLVGLAEQEVNCGMTDDVDVDMVYRELVALRE
SRENLAADKDEVMKQTAKIDSLAEKAKQVWREPRAEDVGSSELAMQLTISLEVGIRS
QVSRKLDLPDOLNAHSEAHGISHQAKMVGTLNIIILDCQELRESYALPTAETDNVPE
WLGTEGGEDESGND"
/complement(158676..158685)
/feature="NMA1222"
/label=DUS
/complement(159354..159357)
/complement(159391..159681)
/feature="NMA1223"
/complement(159391..159681)
/feature="NMA1223"
/feature="NMA1223, len: 96 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1223"
/protein_id="CAB84483.1"
/db_xref="GI:7379912"
/translation="MAESKRVQRLRVFIALDEHPITGLSNKDLVSGLGLTPSQVSRD
IDDLVASGLVILKENGAYAYGIRTKLQIAERPROQHERLQSKIAEIGKRVVD"
/complement(159688..159693)
/complement(159693..159764)
/feature="NMA1224"
/complement(159693..159764)
/feature="NMA1224"
/feature="NMA1224, len: 23 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1224"
/protein_id="CAB84484.1"
/db_xref="GI:7379913"
/db_xref="SPTREMBL:Q9JUN1"
/translation="MVGGRKNGYKGHEIAKKILGLK"
/159975..159978
/159986..162315
/feature="possible partial prophage element pnm5"
/label=pnm5
/159986..160363
/feature="NMA1225"
/159986..160363
/feature="NMA1225"
/feature="NMA1225, possible phage repressor, len: 125 aa;
shows weak similarity to SW:RPC_BPPH1 (EMBL:X02799),
Bacteriophage phi-105 immunity repressor protein (144 aa),
fasta scores; E(): 0.0068, 36.3% identity in 80 aa overlap
and to putative phage repressors. Contains Pfam match to
entry PF01381 HTH_3, Helix-turn-helix and probable
helix-turn-helix motif at aa 21-42 (Score 1857, +5.51 SD).
Lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative phage repressor"
/protein_id="CAB84485.1"
/db_xref="GI:7379914"
/db_xref="SPTREMBL:Q9JUN0"
/translation="MISKSLFGRNLKRRKFLGLTQAAAEKAGIERETWGYERGVSF
MPSGDVLFLNNGIDVSLFAAEQGRPSENSEISESELGYRQASDNGKFFVL
SVAKGAEKAAENTQVSATHKTA"
/160016..160171
/feature="NMA1225"
/feature="NMA1225"
/feature="Pfam match to entry PF01381 HTH_3,
Helix-turn-helix, score 29.50, E-value 7.8e-05"
/160201..160210
/feature="NMA1225"
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/160380..160383
/160392..160637
/feature="NMA1226"

```

```

CDS
160392..160637
/feature="NMA1226"
/feature="NMA1226, len: 81 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1226"
/protein_id="CAB84486.1"
/db_xref="GI:7379915"
/db_xref="SPTREMBL:Q9JUN9"
/translation="MFDPEQOIKWGERAEIIVKEAATONNIEIPPLASALAKAVKVH
YLSQAGVFSFLVEAYATVNPTEKEVDYQAIKEJFEK"
160638..160670
/feature="NMA1227"
/feature="NMA1227, len: 30 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1227"
/protein_id="CAB84487.1"
/db_xref="GI:7379916"
/db_xref="SPTREMBL:Q9JUN8"
/translation="MKRPECILKRIKNYPDIQVIGVVQVLKDLN"
160817..160826
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
160859..160862
/feature="NMA1228"
/feature="NMA1228, len: 322 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1228"
/protein_id="CAB84488.1"
/db_xref="GI:7379917"
/db_xref="SPTREMBL:Q9JUN7"
/translation="MAKTKTTHYLRAQQNHSLFDEVALRRVLSAAPTVDQTQIEQ
YNLTQIMHRTLPDKVKGNSIGGLLIHGSGIKDEHRTMSNRPVQDDHGGTQAPP
SGYSFLRKEAFLVGHVIFCGHGFSLASTVASVLSLSKLNRENPDIVQNIIEFK
AVGNCDKLSLQOQGVKSLILLDSAYQLSNRLYQDSRSTIAKALGKVGVSFTSELS
EELEAQSEIHINLEVLNNGSRASIEAQSLMQEAEIIDDETVMNQGFSITTOQGEVI
QPSDKLSKSVIRIGRYDEANSLPLDSAFATSEYFLEQSRNLTEQ"
161836..162318
/feature="NMA1229"
/feature="NMA1229"
/feature="NMA1229, possible membrane protein, len: 160 aa;
unknown, contains a possible N-terminal signal sequence
and hydrophobic, possible membrane-spanning regions. Lies
within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84489.1"
/db_xref="GI:7379918"
/db_xref="SPTREMBL:Q9JUN6"
/translation="MKKYLPLLRFLLIACISAFIAWKQGFVHGNEKAYDLINVFAL
LACFLIAIMTFLFSDMFEDEDANRQIQIREGVQEQRYIKHSLLFYTYLAVLCVFIVI
LIAHKEYKNGPAIFWLEKSYLFLACISIFYSVFLPGLNLIKSRKEEFKLMKKPKPI

```


misc_feature

domains., score 8.70, E-value 0.28"
149882. .149917
/gene="NMA1206"

/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature"
151105. .151145

/note="Stem loop containing DNA uptake sequences: gaaat
gccgtctgaa ccgcgaagcc ttcagagc attgt"
151110. .151119

misc_feature

/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

misc_feature

complement(151131. .151140)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

gene

151170. .152171

CDS

/gene="NMA1207"
151170. .152171

/note="NMA1207", len: 333 aa; similar to many hypothetical

proteins e.g. TR:O68273 (EMBL:AF026544) Alcaligenes
eutrophus hypothetical protein (322 aa), fasta scores;

E(): 0, 47.3% identity in 315 aa overlap. Also similar to
SW:NIR_RHOCA (EMBL:X72382), nifR3, Rhodobacter capsulatus

nitrogen regulation protein (not stringently required for
nitrogen fixation) (324 aa), fasta scores; E(): 8.1e-19,

31.5% identity in 324 aa overlap. Similar to NMA1633,
fasta scores; E(): 8.3e-20, 34.0% identity in 244 aa

overlap. Contains Pfam match to entry PF01207 UPF0034,
Uncharacterized protein family UPF0034 and PS01136

Uncharacterized protein family UPF0034 signature"
/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1207"

/protein_id="CAB84488.1"

/db_xref="GI:7379898"

/db_xref="SPTREMBL:Q9JUP6"

/translation="MIDQOTTEPKQKTRIIILAPMQLGVDDVMRLLTRIGGYDECVSE
FVRIHTVHGRATWIKYVPEITANGKTRFSGTPCVQLLGSADNMAANALEAVREGAD

KIDLNFCCGAPTVNHRKGGAILLKEPELPHVKTLRERLPAHPIPLAKRRLGYEDKS
PALECAIAEGGACGLTIVARIKAEYEPAPHWIRKINDSDVIPVTANGDVFSLQ

DYIGKITSQNSVLMGRGAVIRPDLARQIYENGPGVKDTPFAEYSKWIQFFELC
LTKEANKYPIARLKLQWLGMKKTFDPAQTLFDRVTVKDADEVRRILNFAFEHINV"

misc_feature

/gene="NMA1207"

151215. .152153

/note="Pfam match to entry PF01207 UPF0034,
Uncharacterized protein family UPF0034, score 249.90,

E-value 3.5e-71"

151479. .151535

misc_feature

/gene="NMA1207"

/note="PS01136 Uncharacterized protein family UPF0034
signature"

complement(152221. .152403)

gene

/gene="NMA1208"

CDS

complement(152221. .152403)

/gene="NMA1208"

/note="NMA1208, len: 60 aa; unknown"

/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1208"

/protein_id="CAB84469.1"

/db_xref="GI:7379899"

/db_xref="SPTREMBL:Q9JUP5"

/translation="MEFWYIGFCFAVVALLSLVNVNAGAFGMQEDDTPQSDYERRLGIG
AKLKNKNTPKSGEKQ"

complement(152408. .152412)

RBS

152471. .152626

/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 212.79"

gene

/label=Correia

complement(153062. .153154)

CDS

/gene="NMA1209"

complement(153062. .153154)

/gene="NMA1209"

misc_feature

complement(153062. .153154)

CDS

/gene="NMA1209"

misc_feature

complement(153062. .153154)

CDS

/gene="NMA1209"

misc_feature

complement(153062. .153154)

CDS

/gene="NMA1209"

/note="NMA1209, len: 30 aa; unknown"
/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1209"

/protein_id="CAB84470.1"

/db_xref="GI:7379900"

/db_xref="SPTREMBL:Q9JUP4"

/translation="MQLFCHAVKYPHDKIMQNWDTGARLHKPV"

complement(153117. .153425)

/gene="NMA1210"

complement(153117. .153425)

/gene="NMA1210"

/note="NMA1210, len: 102 aa; unknown"

/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1210"

/protein_id="CAB84471.1"

/db_xref="GI:7379901"

/db_xref="SPTREMBL:Q9JUP3"

/translation="MRKDRKWTQISAALYQGNKTRSAEAVTRSPKKNPAVNGKSKGAG
AGYGGHURAIQGFYRAGTLPRTVTAQIKREHLKKQARLEESDRACFATVLSCHREISA
"

153153. .153192

complement(153477. .153992)

/gene="NMA1211"

complement(153477. .153992)

/gene="NMA1211"

/note="NMA1211, len: 171 aa; unknown"

/codon_start=1

/transl_table=11

/product="hypothetical protein NMA1211"

/protein_id="CAB84472.1"

/db_xref="GI:7379902"

/db_xref="SPTREMBL:Q9JUP2"

/translation="MVVVTPTQHQTESNGIAIKTVHRRTSFDSRATAGKKSREF
VESCANDKTVELPDEDGKIGVVAEPGSGKGTGIGKARELYAPTACGAGPIADEIA
PDKAFHVTALSAAGLGSVPVRLPYTPPHRSACRRDGRARIAMAGYRHRHTVFKHP
QRNAVPRKKPL"

complement(154017. .154822)

/gene="NMA1212"

/note="possible IS element remnant"

/label=IS

complement(154017. .154822)

/gene="NMA1212"

complement(154017. .154822)

/gene="NMA1212"

/note="NMA1212, pseudogene, possible transposase, len: 162
aa; similar to parts of many putative transposases e.g.
TR:CAB44496 (EMBL:AJ239004) Streptococcus pneumoniae

putative transposase (107 aa), fasta scores; E(): 6.9e-22,
57.9% identity in 107 aa overlap. Highly similar to
NMA1992, fasta scores; E(): 0, 99.0% identity in 103 aa

overlap and NMA1993, fasta scores; E(): 0, 96.3% identity
in 162 aa overlap"

/codon_start=1

/pseudo

/transl_table=11

/product="putative transposase (pseudogene)"

complement(154821. .154862)

/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 53.74"

/label=Correia

complement(154864. .154926)

/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 78.71"

/label=Correia

155009. .155125

/gene="NMA1214"

155009. .155125

/gene="NMA1214"

/note="NMA1214, len: 38 aa; unknown"

/codon_start=1

/transl_table=11

```

/misc_feature
/label=DUS
complement(144347..144356)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(144378..144749)
/gene="NMA1204"
complement(144378..144749)
/gene="NMA1204"
/note="NMA1204"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1204"
/protein_id="CAB84465.1"
/db_xref="GI:7379895"
/db_xref="SPTREMBL:O9J024"
/translations="MKPKIYTHPALPPOAKVGKWSWPGLLFGTLWACFKRMGLGLG
LTGAIFVLVFAQLVGDTPATDSAFNVLGLAVSWFGAGNSLIYARHLLSRGYTELP
ETVEANPQALAOYFGRGGR"
complement(144745..144805)
/note="REP 2; hmms hit to HMM REP 2 (69 - 128), score:
48.94"
complement(144756..144759)
complement(144803..144845)
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 47.30"
/label=Correia
complement(144847..144909)
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 73.39"
/label=Correia
/label=Correia
144981..145022
/note="Stem loop containing DNA uptake sequences: acaat
gccgtctgaa cggcgctttgcc ttacagcgc attgc"
144986..144995
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(145008..145017)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(145044..146735)
/gene="did"
complement(145044..146735)
/gene="did"
/EC_number="1.1.1.28"
/note="NMA1205, dld, probable D-lactate dehydrogenase,
len: 563 aa; similar to e.g. SW:DLD_ECOLI (EMBL:MI0038),
dld, Escherichia coli D-lactate dehydrogenase (EC
1.1.1.28) (570 aa), fasta scores; E(): 0, 60.9% identity
in 565 aa overlap. Contains Pfam match to entry PF01565
FAD_binding_4, FAD binding domain"
/codon_start=1
/transl_table=11
/product="putative D-lactate dehydrogenase"
/protein_id="CAB84466.1"
/db_xref="GI:7379896"
/db_xref="SPTREMBL:O9Jup8"
/translations="MSASOLLRLTQTCGEKYIITDPAKTEQYRQGYRGEGRALAV
RPTGLEMKILQACVEADIVITQAANTGCGSTPDGNDYDRDIVVNTMRNLIQ
TINNNOVCLPGSTLNQLELLKPLGRSPHSVIGSSCIGASVLGGVNCNNGALVOR
GPAYTEMALFAQINEEGRLELVNLGIDLDGTPTEILTNLQGHYQNKDITKQDAGKH
DHAYCEHVQDEPTAARFNADPARHVEASGCAGKLMFAVLDTFPQEKQAVFYIG
TNDINELTDLRRALAGEFSLPVSGEYIHRHAFDIADYVGDYFYVKKFGTHQLPKL
FDLKVARDREGKASPLPKHFSKAMQFYSKYLDPHLPKSMRDYRDKYEHLLIKTGG
KGVDEARAEFKYFAHHSAGFAFECNAETQAMLRFAVASAIRAVHDDREVDLV
ALDIALRRDRDWFELKPSIEDNKIIHKLYYGHFMCHVFHQDYIIKKGNDCMALEHEM
LHLLDQGAQYPAEHNVLHVEAKPALKFYRKLDPTNSFNPGVGKTSKKKNAAE"
complement(146064..146714)
/gene="dld"
/note="Pfam match to entry PF01565 FAD_binding_4, FAD
binding domain, score 153.40, E-value 3.8e-42"
146821..146855
/note="Stem loop containing DNA uptake sequences: aaaa
gccgtctgaa cggct ttacagcgc atcgg"

```

```

misc_feature
146826..146835
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(146841..146850)
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(147217..147226)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
147260..151093
/gene="NMA1206"
147260..151093
/gene="NMA1206"
/note="NMA1206, probable oxidoreductase, len: 1277 aa; the
N-terminal half shows weak similarity to several putative
oxidoreductases e.g. TR:Q9ZLW7 (EMBL:AE001480), gldC,
Helicobacter pylori J99 putative glycolate oxidase (459
aa), fasta scores; E(): 1.5e-09, 24.8% identity in 455 aa
overlap. The C-terminal half shows weak similarity to
putative iron-sulphur proteins e.g. TR:O67558
(EMBL:AE000748), glpC, Aquifex aeolicus oxido/reductase
iron sulfur protein (395 aa), fasta scores; E(): 5.2e-07,
22.8% identity in 465 aa overlap. Contains pfam matches to
entry PF01565 FAD_binding_4, FAD binding domain and to
entry PF00037 fer4, 4Fe-4S ferredoxins and related
iron-sulfur cluster binding domains. Contains PS00198
4Fe-4S ferredoxins, iron-sulfur binding region signature"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAB84467.1"
/db_xref="GI:7379897"
/db_xref="SPTREMBL:O9Jup7"
/translations="MTTITAPQRIEIPYNTSYTDRIVRLIGDRAWLIQDLRQO
RKTGSRARMLFVLGDIWVVRNRPYLVDLLEHPKRRALVRMRHRLNRIKRRDDN
RQVDVYVAAAEKAEVERFSSQKRKHQLERLSKITPHNIMFDGLARVTVTDA
TDMRVEYFVNVNPDTEAEVAPLVRALIELDLVITPRGGTGYTGGAIPLDAMSAVIN
TEKLDKRGVEFVELAGLDGKHPIIRCGAGVTVRVEETAHQAGLVFADPTSDASC
VGVNANNAGKKKAYLWGTALDNLAYWQVNPQGEWLRIERYRHNFKGIHDEETAVFD
VHTLDSGGINIVKTERLEIPGHRFKVGLGKDVDFSLGPGVKQKGGTGIITGVAF
VLHKPKYRTVCMVEFFGTVATATPSIVEIRDFLLAHDVSLRAGLEHLDWRYRAGV
ATKAAGKGRPKMLADVAADDEAAEAAEHICELARADGEGFIAVSPEARKTFWL
DRSRTAAIAKHTNFKINEDWVITPLERLGYSDGIERINIELSTONKLIKLAEOYL
SGKLIDKMGTDLTAFELGERGKHALAHVSAVKARWELLANDLADLADYKARYGSA
VHAAPKAKNESCFATRFDRFLRVSKADVMKPLSEFSKTDTKIILQGLGKIHAKTV
RSRVFALHMHAGGQNVHTNIPVNSDDAEMQLTAIRSVRIMKTARSLLGGVIGEHGI
GRTKLFEFLSDEEMQPFWDYKNQDPKHTFNRLKMGSLDRNATYPSFELLAGESLIM
EKSDLGTIADSVKDCLCRCCKPCVCTHVPANLLYSRPNKILGVLLTFAFLYEQT
RRGVSITKHFEELMDIGDCHCTVCHRCVKPCPVNIDFGDVTAVRNYLADSGHKRFAPA
AMGAFINATGPKTIKALRAAMTGTGPAONFAVYKIGKLLPIGTPKKOAEKPTAVGKA
PIKEQVTHFINRPLPKSVPAKTRSLSLGIEDGKSIPIIRNPAKAEASTNRRVAFHRM
ERLFSQILAGLAWLHVGVQTVLPFGMCCGIPQDAGGNKAKAEASTNRRVAFHRM
ANTLNLDITKVYVSCGTCYDQLEKRYFEIFFGCRIIDTHEYLKGVKLVKVGKQO
YLYHDPCHTPIKTMNATQMASSLMGQKVLSDRCGSGFMFAVRPDIAIQVKFRKE
EIEKNLKEPQGEPPVKMLTSCPACLGSLRSYTDNNNPADYIVIEMAKHILGENWLDE
FVKKANNGVKVVLL"
147662..148426
/gene="NMA1206"
/note="Pfam match to entry PF01565 FAD_binding_4, FAD
binding domain, score 73.30, E-value 5.2e-18"
complement(148276..148285)
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(148876..148885)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(149397..149406)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
149666..149752
/gene="NMA1206"
/misc_feature
/note="Pfam match to entry PF00037 fer4, 4Fe-4S
ferredoxins and related iron-sulfur cluster binding

```


| | | | | | | | | | | | | |
|-------------|---|-------------|--|-------------|-----------|--------------|-----|------|-----|--------------|-----------|--------------|
| repeat_unit | 140918..140978 /note="92% identity to consensus CTAGGACGTAAATCTAAAGAACCGTTTACCCGATAAGTTTCGGACCGACAGAC CTAG" /label=RS15 140979..140998 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 complement(140999..141095) /note="98% identity to consensus GACGAGATTTAGATTATGGAAATTCAGGAATGATTGAATCATTAGAAAAACCA CAGNATCTATCAGAAAAACAGAAACCCACCGCCGTC" /label=RS53 141096..141115 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 141116..141176 CTAGGACGTAAATCTAAAGAACCGTTTACCCGATAAGTTTCGGACCGACAGAC CTAG" /label=RS15 141177..141196 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 141198..141253 /note="81% identity to consensus GACGTCGGGGTTCGTCTTTTCCGATAGATCCCGTGGTTTTCGGTTACTGG" 141254..141273 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 complement(141274..141330) /note="98% identity to consensus CCAGCAACACAGCAGCAGACTTTATCGGAAAAACAGAAACCCACCGACGTC" /label=RS42 141331..141350 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" 141488..141498 /note="(g)11" complement(141597..141751) /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 277.40" /label=Correia complement(141821..141991) /gene="NMA1201" complement(141821..141991) /gene="NMA1201" /note="NMA1201, probable rubredoxin, len: 56 aa; similar to many e.g. SW:RUBR_ACICA (EMBL:246863) Acinetobacter calcoaceticus rubredoxin involved in the hydrocarbon hydroxylating system (54 aa), fasta scores; E(): 1.6e-13, 64.2% identity in 53 aa overlap. Contains Pfam match to entry PF00301 rubredoxin, Rubredoxins and PS00202 Rubredoxin signature" /codon_start=1 /transl_table=11 /product="putative rubredoxin" /protein_id="CAB84462.1" /db_xref="GI:7379892" /db_xref="SPTREMBL:Q9JQN3" /translacion="MAQYMGPGCWYDERHGDPEHGIAPTKFDIPDDWKCEGV KREDVILLDFVI" complement(141845..141991) /gene="NMA1201" /note="Pfam match to entry PF00301 rubredoxin, Rubredoxins, score 97.40, E-value 1.1e-26" complement(141863..141895) /gene="NMA1201" /note="PS00202 Rubredoxin signature" complement(141999..142002) complement(142012..142091) /note="Correia element; hmms hit to HMM Correia (76 - | repeat_unit | 156), score: 105.49" /label=Correia complement(142092..142152) /note="Correia element; hmms hit to HMM Correia (1 - 61), score: 78.03" /label=Correia 142244..142263 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 complement(142264..142320) /note="RS101" /label=RS101 142321..142340 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 complement(142341..142395) /note="RS102" /label=RS102 142396..142415 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" /label=DRS3 complement(142449..143540) /gene="NMA1202" complement(142449..143540) /gene="NMA1202" /note="NMA1202, probable acyl-CoA dehydrogenase, len: 363 aa; shows weak similarity to many e.g. SW:IVD_HUMAN (EMBL:M34192), 1vd, Homo sapiens isovaleryl-CoA dehydrogenase precursor (EC 1.3.99.10) (423 aa), fasta scores; E(): 2.7e-10, 24.0% identity in 358 aa overlap. Contains Pfam match to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase" /codon_start=1 /transl_table=11 /product="putative acyl-CoA dehydrogenase" /protein_id="CAB84463.1" /db_xref="GI:7379893" /db_xref="SPTREMBL:Q9JU00" /translation="MNAQTLIANVAEFVKTKLPVDDIDRKYGYPEAFMRELGAIGG FGAVTEAGGNGGLATQIATVREIGKECATSFSAWCAACAWYLHQTNPRAVKOK YLADILQKVLGAGTMSNTVKHLADIEKHNLQAEVGGYTVNGALPWNSIGEDHW ANTAOIGGYVMFTTGGWEGVSLQNGPEFCALERTFSLNFKDVFIPDSDIIAAPE QFADYIOSIKAGTLLQIGIGAGVIDSLGIIRLANVVAEVNRYLDDGDVLSLRD GAWAETERLAGWSGTPDNLTLKREAAVLTAAQAALHSGAGKYLMRSPAQR RVREAMEVAIVTPAIKHLRKEIAAEAAK" complement(143184..143513) /gene="NMA1202" /note="Pfam match to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase, score 21.30, E-value 2e-05" complement(143547..143550) complement(143675..144223) /gene="NMA1203" complement(143675..144223) /gene="NMA1203", len: 182 aa; shows weak similarity to TR.Q53427 (EMBL:S71704), mip, Legionella pneumophila macrophage infectivity potentiator protein (184 aa), fasta scores; E(): 2.9e-08, 26.9% identity in 175 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1203" /protein_id="CAB84464.1" /db_xref="GI:7379894" /db_xref="SPTREMBL:Q9JUP9" /translation="MARLTVHTLETAPEAAKARVAVLQNNGFTPLNIGVLSNAPEAL AFYOEVGKLANSLTAGEVEVIQIIAARTNOCGFCVAGHTKLATLKKLSEOSYKAA RALAAGEFDADKALGAALAAQTQAVMAKKGAVSDEELKAFFDAGYNQQQAVVVMGVALA TLCNVNNLQTEINPELQAIYA" 144327..144360 /note="Stem loop containing DNA uptake sequences: gaaat gccgtctgaa cattt ttcagacgac acga" 144332..144341 /note="Core DNA uptake sequence: gccgtctgaa" | repeat_unit | stem_loop | misc_feature | RBS | gene | CDS | misc_feature | stem_loop | misc_feature |
|-------------|---|-------------|--|-------------|-----------|--------------|-----|------|-----|--------------|-----------|--------------|


```

/feature="NM1189"
/feature="NM1189", len: 54 aa; unknown, identical to NM1304
and NM1863"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1189"
/protein_id="CAB84450.1"
/db_xref="GI:7379880"
/db_xref="SPTREMBL:Q9JRV3"
/translation="MGQVAFYEKIMIGLSAKSRSEASEQADLAEPFAEGELANYQEML
KRHLQTKSVE"
134118..134354
/feature="NM1190"
134118..134354
/feature="NM1190"
/feature="NM1190", len: 78 aa; unknown, identical to NM1305
and NM1862"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1190"
/protein_id="CAB84451.1"
/db_xref="GI:7379881"
/db_xref="SPTREMBL:Q9JQF9"
/translation="MRILDFKNPATGNVSHSLWANVACAAAGTVKFMPLDPESAETW
AVYLGIVGVAVARSLVSVKQEVNESRETAGE"
134314..134319
/feature="NM1190"
134326..134745
/feature="NM1191"
134326..134745
/feature="NM1191"
/feature="NM1191", len: 139 aa; unknown, highly similar to
NM1306, fasta scores; E(): 0, 97.1% identity in 139 aa
overlap and NM1861, fasta scores; E(): 0, 97.1% identity
in 139 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1191"
/protein_id="CAB84452.1"
/db_xref="GI:7379882"
/db_xref="SPTREMBL:Q9JUQ5"
/translation="MNLVKLLANNWQPIAIIALVGTGLAVSHHQYKSAFAKQAVIE
KMKRDKAQLLSAQNARELEQARAEAKYEVKAHVAGMALAKQAEVSRKTENKK
ELENVLTQRKNAGGCCIDGFGHGLQLYKRALGYGN"
134606..134615
/feature="NM1191"
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
134765..134947
/feature="NM1192"
134765..134947
/feature="NM1192"
/feature="NM1192", len: 60 aa; unknown, highly similar to
NM1860, fasta scores; E(): 7.9e-26, 98.3% identity in 60
aa overlap and to part of NM1307, fasta scores; E():
6.8e-09, 82.8% identity in 29 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1192"
/protein_id="CAB84453.1"
/db_xref="GI:7379883"
/db_xref="SPTREMBL:Q9JUQ4"
/translation="MPTPPAALMVAPVRPNPPKDKGTAILLEHAAEFGGVVALENQN
QAWRDWAGHSRKVGN"
complement(134992..135162)
/feature="NM1193"
complement(134992..135162)
/feature="NM1193"
/feature="NM1193", len: 56 aa; unknown, identical to NM1859"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1193"
/protein_id="CAB84454.1"

/feature="NM1194"
/feature="NM1194", len: 72 aa; unknown, identical to NM1858"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1194"
/protein_id="CAB84455.1"
/db_xref="GI:7379885"
/db_xref="SPTREMBL:Q9JCS39"
/translation="MNEYTFSYRENGKSWLSIWADNPPEARAKFRAARENHYDGEV
VAKVYTFVNISWKKLYKRYKYLMIKE"
135390..135728
/feature="NM1195"
135390..135728
/feature="NM1195"
/feature="NM1195", len: 112 aa; unknown, highly similar to
part of NM1308, fasta scores; E(): 2e-20, 84.4% identity
in 64 aa overlap and similar to part of NM1857, fasta
scores; E(): 7.2e-17, 47.3% identity in 112 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1195"
/protein_id="CAB84456.1"
/db_xref="GI:7379886"
/db_xref="SPTREMBL:Q9JRI7"
/translation="MTYRELVERQLAVRHADLELGLSRAREQEPFVHVSDLLDKAGI
EYARMDKDFQTFTHLEYPTNYDTKRAVWQTLGAYVCVNDGDGLGELIASNRPDGVA
VRIVFGDVPV"
135740..136006
/feature="NM1196"
135740..136006
/feature="NM1196"
/feature="NM1196", len: 88 aa; unknown, highly similar to
NM1310, fasta scores; E(): 6.8e-34, 94.3% identity in 88
aa overlap and similar to NM1856, fasta scores; E():
3.5e-11, 45.2% identity in 84 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1196"
/protein_id="CAB84457.1"
/db_xref="GI:7379887"
/db_xref="SPTREMBL:Q9JUQ3"
/translation="MDFEFGFRLWPIATAAFWFWNGISGRLEADKRIDLKKEULH
AVKLSYHTKQDAKADRDNTAAASLGRIENKLEKYNEKLDKADKS"
135819..135828
/feature="NM1196"
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
136003..136200
/feature="NM1197"
136003..136200
/feature="NM1197"
/feature="NM1197", len: 65 aa; unknown, similar to NM1311,
fasta scores; E(): 8.2e-14, 56.3% identity in 64 aa
overlap and NM1855, fasta scores; E(): 3.2e-09, 45.5%
identity in 66 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NM1197"
/protein_id="CAB84458.1"
/db_xref="GI:7379888"
/db_xref="SPTREMBL:Q9JUQ2"
/translation="MSAPILEALARIENKTDQTLKNQKEMQAEIAQIRQDKTRTAITF
GALGGGVITVGWELLKAKMGL"
136203..136709
/feature="NM1198"

```

/db_xref="SPTREMBL:Q9JUQ8"
/translation="MPSIKRALISLSDKGTGAVEFAOTLHKLGVLEILSTGGTAKLLADA
GVPVIVADYTGPEMLDGRVKTLPKIHGGILIGREDLPEHVAKMEEHGIGNDILVCV
NLYPFAATATKNCITLEDIAENIDIGGPTWVRSAAKNWKAHVAVTDPADPFAIAEME
ANGALSDKRETNLSKRAFSTAQYDGMISNLYLTSLSDDVLSGTPPIGEFFPQFNQSW
IKVQDMRYGENPHORAAFYRDVYPAAAGSLAYSAYKQOGKELSYNTIADADAWEAVKSF
DAPACVIVHANPCGVAVAADTLTAYKLAYATDTTSAFGGIIAFNREVVDGETVVKQITD
NOFMEVLMAPKFTAETAEIAAAKNVRVLIQISLTTPLEAGANRFEELKRVGGGLLVOTP
DIYLRNADLVKVSFKPOLTQEWNDLMFYNNVAKYVKSNAIVFGKGGQYVYIGAGOMS
RVDSTRIARKAQDANLDLNGACAASDAFFPRDGVVIAEOGIKAILHPAGSMRDQE
VFDADEHGIAWVVTGVHRFH"
/codon_start=1
/note="Stem loop containing DNA uptake sequences: gtaat
gcgcttgaa ggcct ttcagacgcc atttt"
130339. .130373
130344. .130353
/note="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
/complement(130359. .130368)
/note="Core DNA uptake sequence: gcgctctgaa"
/label=DUS
130452. .130456
130468. .130692
/genes="NMA1183"
130468. .130692
/genes="NMA1183"
/codon_start=1
/note="NMA1183, len: 74 aa; unknown"
/transl_table=11
/product="hypothetical protein NMA1183"
/protein_id="CAB84445.1"
/db_xref="GI:7379875"
/db_xref="SPTREMBL:Q9JUQ7"
/translation="MGCRCVCSPLHOKFLENRRRESAVLSAGTVEMGNIDIGSGGL
KPLTVWNHCIIATNRPATRAHPRHRTTVA"
complement(130798. .130952)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 240.56"
/label=Correia
130937. .131890
/genes="NMA1185"
/note="possible IS element"
130937. .131893
/genes="NMA1185"
130937. .131893
/genes="NMA1185"
/note="NMA1185, probable transposase, len: 318 aa; similar
to many e.g. TR:Q51474 (EMBL:M24471), Alteromonas atlantica
CDS within insertion element IS492 which is involved in
reversible inactivation of extracellular polysaccharide
production (318 aa), fasta scores: E(): 1.3e-28, 33.9%
identity in 301 aa overlap. Also shows weak similarity to
pilin-inverting genes e.g. SW:PIV_MORLA (EMBL:M34367),
piv, Moraxella lacunata pilin gene inverting protein (322
aa), fasta scores: E(): 8.5e-14, 30.4% identity in 313 aa
overlap. Highly similar to others from Neisseria
meningitidis e.g. NMA0733, fasta scores: E(): 0, 94.3%
identity in 279 aa overlap, also NMA2085. Contains Pfam
match to entry PF01548 Transposase_9, Transposase"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="CAB84446.1"
/db_xref="GI:7379876"
/db_xref="SPTREMBL:Q9JUQ6"
/translation="MLIHVIDIAKRNFIJAVSSLSKTTKTETNNPKGIAHTVLYLKKHK
VALVYESTGGLEIIPAARIRAGIAVIITANRPTHQASQPLTKTDAKAKMAFF
AQMKAQEDWQMPYHPPTPEAEVEALVNRNQPADMTAEKRLHWHVETQVESVK
QLIAHFDRLLIDELQOIDHTTHFDGKAQVAEOIKIGISITTTATLMALPELGRISH
KRASIVGTAPHRESGETKFKSCFCGGSAYRKALYMATVAATREPLIRDFYORLP
SGGPKYKAVATACMRKLLTISNARMRDYFAENDTAENGIRTA"
131126. .131431
/genes="NMA1185"
/note="Pfam match to entry PF01548 Transposase_9,

Transposase, score 34.10, E-value 3.3e-06"
131989. .132483
/genes="NMA1186"
131989. .136706
/note="possible partial prophage element pnm3"
/label=pnm3
131989. .132483
/genes="NMA1186"
/note="NMA1186, len: 164 aa; unknown, similar to part of
TR:Q38494 (EMBL:M64097) bacteriophage Mu protein E16 (195
aa), fasta scores: E(): 2.7e-05, 46.7% identity in 60 aa
overlap. Identical to NMA1301 and similar to NMA1867,
fasta scores: E(): 1.5e-25, 55.5% identity in 137 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1186"
/protein_id="CAB84447.1"
/db_xref="GI:7379877"
/db_xref="SPTREMBL:Q9JS31"
/translation="MYFICIGVNWSIYCVYEFEMPRALIAKTIQAQKELGDDGTYR
AVLERTVDARSCADMVSELESVADVDMRSHGPKAKGNPHKPHLRPTSSAAMLQKV
EALLTVGGKHNVYAHAMARRMFGKDKVEYLDLDTQLHLKLVAAUQIAENRRTKRGAGDDG
VRKS"
132464. .132892
/genes="NMA1187"
132464. .132892
/genes="NMA1187"
/note="NMA1187, len: 142 aa; unknown, identical to NMA1302
and similar to NMA1866, fasta scores: E(): 2.4e-10, 29.7%
identity in 145 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1187"
/protein_id="CAB84448.1"
/db_xref="GI:7379878"
/db_xref="SPTREMBL:Q9JRW0"
/translation="MGCFKEVHLDPDTLVDIVDVLGAATEQLVKAIGGARFKFGKGGK
VDFERLAILVEALGVKTHLLOVYGGEELYVPRCKALIOLENRHFYQEFVKLRDID
KKSGLMAMTKLCPKPKYGOISSRTGYTTIINEMSRPAAQQAALF"
complement(133017. .133171)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 283.48"
/label=Correia
133189. .133193
133200. .133745
/genes="NMA1188"
133200. .133745
/genes="NMA1188"
/note="NMA1188, possible amidase, len: 181 aa; shows very
weak similarity to TR:P89923 (EMBL:Y11476), p1y12
Bacteriophage 12826 N-acetylmuramoyl-L-alanine amidase
endolysin (EC 3.5.1.28) (257 aa), fasta scores: E(): 0.16,
26.0% identity in 177 aa overlap and to bacterial amidases
(some putative) e.g. TR:Q32421 (EMBL:D76414), lytH,
Staphylococcus aureus putative N-acetylmuramoyl-L-alanine
amidase (291 aa), fasta scores: E(): 0.0093, 25.8%
identity in 182 aa overlap. Identical to NMA1303 and
highly similar to NMA1865, fasta scores: E(): 0, 96.7%
identity in 181 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="CAB84449.1"
/db_xref="GI:7379879"
/db_xref="SPTREMBL:Q9JR25"
/translation="MGKTVTLTAGHSNTPDPGAVNGSDREADLAQDMRNIVAAI LRDDY
GLTVKTDGCKGNMPLREAVKLI RGSDDVAIEFHTNAAVSKAATGIEALSTVKNKRWQ
VLKAVAKKTKNKL RGEDGFKPDNAGQHSRLAYAQAGGIVFEFFISNDTDLALFKTT
KWGICRAIADATAMEUGAARV"
133951. .134115
/genes="NMA1189"
133951. .134115

misc_feature

/gene="pnta"
/note="Pfam match to entry PF01262 AlaDh_PNT, Alanine dehydrogenase/pyridine nucleotide transhydrogenase, score 697.80, E-value 5.3e-206"
complement(124999..125008)
/gene="pnta"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(125572..125649)
/gene="pnta"
/note="PS00837 Alanine dehydrogenase & pyridine nucleotide transhydrogenase signature 2"
complement(126164..126167)
complement(126284..126400)
/gene="NMA1178"
complement(126284..126400)
/gene="NMA1178"
/note="NMA1178, len: 38 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1178"
/protein_id="CAB84440.1"
/db_xref="GI:7379870"
/db_xref="SPTREMBL:Q9JUR2"
/translation="MLRRAGLQSENFKRRQYVSDNTGKRCNYCQPDENSL"
126446..127279
/gene="serB"
126446..127279
/gene="serB"
/EC_number="3.1.3.3"
/note="NMA1179, serB, probable phosphoserine phosphatase, len: 277 aa; similar, except at N-terminus, to many e.g. SW:SERB_ECOLI (EMBL:X03046), serB, Escherichia coli phosphoserine phosphatase (EC 3.1.3.3) (322 aa, fasta scores; E(): 8.4e-30, 45.7% identity in 234 aa overlap)
/codon_start=1
/transl_table=11
/product="putative phosphoserine phosphatase"
/protein_id="CAB84441.1"
/db_xref="GI:7379871"
/db_xref="SPTREMBL:Q9JUR1"
/translation="MFOALVQPPSAEPALPSPRLPEPCDADEKMRFRIVEGFSL SGDAALLSCQIDHVLDPDMDFGELGIVSDMDTLITIECVDEIAEGVLKNKVAE ITERSNGELFQSLRVSALLAGLDEKILADVENVLKLSPCAEFLDECKRHQVY FLIVSGGFTFFERLOQRLGFEQHANILEIENGRLTGRLKRIIDAQAKADLLREYR SLILQPHQVLANGDCANDIPMLKEAGIGVATRAKPKARAADACINFGLEKRVRLF G"
127038..127047
/gene="serB"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

misc_feature

127292..127328
/note="Stem loop containing DNA uptake sequences: cggat gccgtctgaa aacagtg ttcagacgc atttg"
127297..127306
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(127314..127323)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

stem_loop

127335..127338
127345..127890
/gene="NMA1180"
127345..127890
/gene="NMA1180"

misc_feature

127345..127890
/note="NMA1180, probable integral membrane protein, len: 181 aa; shows weak similarity to the N-terminal half of Escherichia coli hypothetical proteins e.g. TR:P77195 (EMBL:D90802) Escherichia coli hypothetical protein (430 aa), fasta scores; E(): 3.2e-05, 26.8% identity in 183 aa overlap. Also shows weak similarity to the N-terminal half of NMA0434, fasta scores; E(): 0.018, 27.7% identity in 119 aa overlap. The following CDS, NMA1181, is similar to

RBS

gene

CDS

127915..128568
/gene="NMA1181"
/note="NMA1181, probable integral membrane protein, len: 217 aa; shows weak similarity to the C-terminal half of bacterial hypothetical proteins e.g. SW:YADQ_ECOLI (EMBL:D26562), yadQ, Escherichia coli hypothetical protein (436 aa), fasta scores; E(): 2.2e-05, 25.0% identity in 212 aa overlap. Also shows weak similarity to the C-terminal half of NMA0434, fasta scores; E(): 7.2e-06, 25.6% identity in 211 aa overlap. The previous CDS, NMA1180, is similar to the N-termini of these proteins and is separated by a {gcc}4 repeat. Contains hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84443.1"
/db_xref="GI:7379873"
/db_xref="SPTREMBL:Q9JUQ9"
/translation="MLTSVATAVARIGLDGVQOYHPANLAVNTSLNFSVAVIGPILG ATAVYFRRSAKEPFFLRNDRPRIILAIFALIGAIVAFPEITLGNKAGNQLTGG LTDWQHSLELTAVKVLNMLAAGAYGGLTSPMLGSTISFAAAAMNVSFPEMPS ESAAVVGAAYFLGVLSNMLPLTAAVFLVLELYTAPALIMPLCTGMAGAVMTARKMGCK"
128388..128397
/gene="NMA1181"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
128577..128613
/note="Stem loop containing DNA uptake sequences: aaaa gccgtctgaa accggtt ttcagacgc attcc"
128582..128591
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(128599..128608)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
128721..128724
128732..130324
/gene="purH"
128732..130324
/gene="purH"
/EC_number="2.1.2.3"
/EC_number="3.5.4.10"
/note="NMA1182, purH, probable bifunctional purine biosynthesis protein, len: 530 aa; similar to many e.g. SW:PUR9_ECOLI (EMBL:J05126), purH, Escherichia coli bifunctional purine biosynthesis protein (EC 2.1.2.3) (EC 3.5.4.10) (529 aa), fasta scores; E(): 0, 59.5% identity in 528 aa overlap. Probably encodes phosphoribosylaminoimidazolecarboxamide formyltransferase and IMP cyclohydrolase"
/codon_start=1
/transl_table=11
/product="putative bifunctional purine biosynthesis protein"
/protein_id="CAB84444.1"
/db_xref="GI:7379874"

misc_feature

the C-termini of these proteins and is separated by a {gcc}4 repeat. Contains possible N-terminal signal sequence and hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84442.1"
/db_xref="GI:7379872"
/db_xref="SPTREMBL:Q9JUR0"
/translation="MGRKLMFALAAAGVIGGLVIGIVLTLMHFIOHTAYGYGADGVY TSFREGVAQASGMRRVAVLTLCGAIAGGWMLLKRFKGPQIEKAAKOPQLOGIPFLT TVFHVLQIITVGLGSPLGREVAPREMTAFAPAGGRRGLDEDEMRLLIACASGAGL AAYNVPLASTLFILEAMLVG"
127915..128568
/gene="NMA1181"
127915..128568
/gene="NMA1181"
/note="NMA1181, probable integral membrane protein, len: 217 aa; shows weak similarity to the C-terminal half of bacterial hypothetical proteins e.g. SW:YADQ_ECOLI (EMBL:D26562), yadQ, Escherichia coli hypothetical protein (436 aa), fasta scores; E(): 2.2e-05, 25.0% identity in 212 aa overlap. Also shows weak similarity to the C-terminal half of NMA0434, fasta scores; E(): 7.2e-06, 25.6% identity in 211 aa overlap. The previous CDS, NMA1180, is similar to the N-termini of these proteins and is separated by a {gcc}4 repeat. Contains hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84443.1"
/db_xref="GI:7379873"
/db_xref="SPTREMBL:Q9JUQ9"
/translation="MLTSVATAVARIGLDGVQOYHPANLAVNTSLNFSVAVIGPILG ATAVYFRRSAKEPFFLRNDRPRIILAIFALIGAIVAFPEITLGNKAGNQLTGG LTDWQHSLELTAVKVLNMLAAGAYGGLTSPMLGSTISFAAAAMNVSFPEMPS ESAAVVGAAYFLGVLSNMLPLTAAVFLVLELYTAPALIMPLCTGMAGAVMTARKMGCK"
128388..128397
/gene="NMA1181"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
128577..128613
/note="Stem loop containing DNA uptake sequences: aaaa gccgtctgaa accggtt ttcagacgc attcc"
128582..128591
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(128599..128608)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
128721..128724
128732..130324
/gene="purH"
128732..130324
/gene="purH"
/EC_number="2.1.2.3"
/EC_number="3.5.4.10"
/note="NMA1182, purH, probable bifunctional purine biosynthesis protein, len: 530 aa; similar to many e.g. SW:PUR9_ECOLI (EMBL:J05126), purH, Escherichia coli bifunctional purine biosynthesis protein (EC 2.1.2.3) (EC 3.5.4.10) (529 aa), fasta scores; E(): 0, 59.5% identity in 528 aa overlap. Probably encodes phosphoribosylaminoimidazolecarboxamide formyltransferase and IMP cyclohydrolase"
/codon_start=1
/transl_table=11
/product="putative bifunctional purine biosynthesis protein"
/protein_id="CAB84444.1"
/db_xref="GI:7379874"

gene

CDS


```

/db_xref="SPTREMBL:Q9JUR20"
/translation="MEDEDFIRAAASMLAEVEVFGFIYKESKSGGGRFYERCWLMGSED
VLYGRVHFGGQNNITLFLTGTGGVAKGWSERLFAELTNAIRPKITFVDIAKDFFN
GEISPNQAREDRNKGMTCHHVKPKGCLGSDWEDEDAKTKTKTYIGKRESKYY
RVYKGLGDKTSTWTFEFIEFNKADIVIFEVLQNGEYFGGAYPICERFVQKATR
IHAKEDVISADRIEVLWKKQFGRAANGLKFIPELDAKLKELIEPSHHKLPKSLA
PEAYDCAPLKAQA THEQPAFKPKDPYNYEYENLEKQLEQOKHVNNEESYNNFIYD
KFARLPIPSWA"
118523..121495
/notes="region containing CDS similar to NMA0776-NMA0782,
NMA1792-NMA1797 and NMA1999-NMA2005"
118943..118966
/genes="NMA1167"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
119548..119892
/genes="NMA1168"
119548..119892
/genes="NMA1168"
/notes="NMA1168, len: 114 aa; unknown, similar to others
from Neisseria meningitidis, NMA2000, fasta scores; E():
0, 100.0% identity in 114 aa overlap, NMA1793, fasta
scores; E(): 1.6e-13, 37.8% identity in 98 aa overlap and
NMA0781, fasta scores; E(): 2.3e-10, 34.9% identity in 83
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1168"
/protein_id="CAB84430.1"
/db_xref="GI:7379860"
/db_xref="SPTREMBL:Q9JS46"
/translation="MGLKCLPERLITQGNKMNIOQLGHIVGVKFKNGIEGRSFYD
CRLIVATPLDSSQGNALGSSSTEYDFGGSANFEQFNAQPIEANLVEIITCKTKQ
LKVIGFQPVKKG"
119897..120094
/genes="NMA1169"
119897..120094
/genes="NMA1169"
/notes="NMA1169, len: 65 aa; unknown, similar to others
from Neisseria meningitidis, NMA2001, fasta scores; E():
4.2e-27, 100.0% identity in 65 aa overlap, NMA1794, fasta
scores; E(): 4.6e-07, 38.3% identity in 60 aa overlap and
NMA0780, fasta scores; E(): 1.3e-06, 36.1% identity in 61
aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1169"
/protein_id="CAB84431.1"
/db_xref="GI:7379861"
/db_xref="SPTREMBL:Q9JQP6"
/translation="MOKVYVQSVSTGDFLYLSPETGDIGHTKLTITNADYFYDFEAI
NAGLEIGNQYEFVWFGLKD"
120102..120386
/genes="NMA1170"
120102..120386
/genes="NMA1170"
/notes="NMA1170, possible integral membrane protein, len:
94 aa; unknown, similar to others from Neisseria
meningitidis, NMA2002, fasta scores; E(): 0, 100.0%
identity in 94 aa overlap. Contains hydrophobic, possible
membrane-spanning regions, NMA1170, fasta scores; E(): 0,
100.0% identity in 94 aa overlap and NMA0779, fasta
scores; E(): 1.5e-18, 72.2% identity in 72 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84432.1"
/db_xref="GI:7379862"
/db_xref="SPTREMBL:Q9JRX7"
/translation="MFGRLKNAPSITAKHFLKENTMKFINTCRKYGAKLAVVTAAPL
ALAAHANATLPATKNALEAAKADGMEAGNIVVGFAALFVSIVKRVMK"
120393..120671
/genes="NMA1171"
120393..120671

```

```

/genes="NMA1171"
/notes="NMA1171, len: 92 aa; unknown, similar to others
from Neisseria meningitidis, NMA2003, fasta scores; E():
0, 100.0% identity in 92 aa overlap, NMA0778, fasta
scores; E(): 1.5e-25, 58.7% identity in 92 aa overlap and
NMA1796, fasta scores; E(): 8.4e-09, 32.3% identity in 93
aa overlap. C-terminus is hydrophobic"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1171"
/protein_id="CAB84433.1"
/db_xref="GI:7379863"
/db_xref="SPTREMBL:Q9JRF3"
/translation="MYQVGNKCLEKHOAENLYFLSVVPRIKENGQIVRPEYNGSLWK
MSDQPLRLLEACEPSKDNQSLGLETGWIVFGILASVYFVSLKKVLK"
120799..121116
/genes="NMA1172"
120799..121116
/genes="NMA1172"
/notes="NMA1172, len: 105 aa; similar to others from
Neisseria meningitidis e.g. TR:087782 (EMBL:AJ010115)
Neisseria meningitidis strain B:15:P1.16 hypothetical
protein (fragment) (148 aa), fasta scores; E(): 0, 83.8%
identity in 105 aa overlap. Also similar to NMA2004, fasta
scores; E(): 4e-25 54.8% identity in 104 aa overlap and
NMA0777, fasta scores; E(): 6.7e-16, 57.5% identity in 73
aa overlap. Lies within a region of unusually low GC
content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1172"
/protein_id="CAB84434.1"
/db_xref="GI:7379864"
/db_xref="SPTREMBL:Q9JUR8"
/translation="MNKPFITQAOALVLYKQYPSKYGQSMALIAQKEFEFVNKVE
YDILESFSYFLNKRVAHNWKIYFSDSESVIFIRKEENGKTVHEFVQYQYFDDSDPNS
MFE"
121058..121498
/genes="NMA1173"
121058..121498
/genes="NMA1173"
/notes="NMA1173, probable cell-surface protein, len: 146
aa; highly similar to the N-terminus of TR:087783
(EMBL:AJ010115), tspp, Neisseria meningitidis strain
B:15:P1.16 T-cell and B-cell antigen (partial CDS) (504
aa), fasta scores; E(): 5.9e-29, 68.2% identity in 129 aa
overlap. Similar to others from Neisseria meningitidis,
e.g. NMA0776, fasta scores; E(): 1.8e-23, 51.0% identity
in 145 aa overlap, and NMA1797, fasta scores; E():
1.8e-23, 51.0% identity in 145 aa overlap (both tspp
genes) and NMA2005 (another putative cell-surface
protein), fasta scores; E(): 1.2e-30, 66.2% identity in
133 aa overlap. N-terminal half has low GC content.
Contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative cell-surface protein"
/protein_id="CAB84435.1"
/db_xref="GI:7379865"
/db_xref="SPTREMBL:Q9JUR7"
/translation="MNSFKSTFIVLILICSSNSYSEPVRLERSQIKFOSSNNLKSS
GFKLDSSSKSFARFIEAENFHIPTGAKARINAKITASVRAGVLAVGVLARLGAKF
STRAPVYGTALLAHDIYETFKEDIQARGYQYDPETDKPAKVSQ"
121676..121694
/notes=">= 85% match to ATTCCCNACNCTTTTCGTC"
/label=RS3x
121695..121714
/notes=">= 90% match to ATTCCCNNNNNNNNGGAAT"
/label=drs3
121715..121770
/notes="100% identity to consensus
CTAGAATCTCGGACTTTCAGATAATCTTTGAATATTGCTGTGTTCTTAAGGTCTAG"
/label=RS13

```



```

/note="98% identity to consensus
CAAATTCGTCGGTTTCAGCGGTTTCGATAAATCACTGTAACGTTAAGTTCCCGG"
/label=RS41
112072..112091
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112092..112142)
/note="98% identity to consensus
GACGGCGTAAAGGTTCTGTGTTTTTAAGTTTAGGTAACTAATGAGTCGTC"
/label=RS62
112143..112162
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112163..112218)
/note="100% identity to consensus
CAAATTCGTCGGTTTCAGCGGTTTCGATAAATCACTGTAACGTTAAGTTCCCGG"
/label=RS41
112219..112238
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112239..112296)
/note="RS100"
/label=RS100
112297..112316
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112317..112372)
/note="81% identity to consensus
CCAGTCGGTCAGTTTCAGTCATTTCCGATAAATTCCTGTGCTTTTCATTCTAG"
/label=RS17
112373..112392
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112393..112443)
/note="98% identity to consensus
GACGGCGTAAAGGTTCTGTGTTTTTAAGTTTAGGTAACTAATGAGTCGTC"
/label=RS62
112444..112463
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112464..112519)
/note="100% identity to consensus
CAAATTCGTCGGTTTCAGCGGTTTCGATAAATCACTGTAACGTTAAGTTCCCGG"
/label=RS41
112520..112539
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112540..112590)
/note="99% identity to consensus
GACGGCGTAAAGGTTCTGTGTTTTTAAGTTTAGGTAACTAATGAGTCGTC"
/label=RS62
112591..112610
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112611..112666)
/note="100% identity to consensus
CAAATTCGTCGGTTTCAGCGGTTTCGATAAATCACTGTAACGTTAAGTTCCCGG"
/label=RS41
112667..112686
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112687..112737)
/note="99% identity to consensus
GACGGCGTAAAGGTTCTGTGTTTTTAAGTTTAGGTAACTAATGAGTCGTC"
/label=RS62
112738..112757
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112758..112813)
/note="82% identity to consensus
CCAGTCGGTCAGTTTCAGTCATTTCCGATAAATTCCTGTGCTTTTCATTCTAG"
/label=RS17
112814..112833

```

```

/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112834..112884)
/note="100% identity to consensus
GACGGCGTAAAGGTTCTGTGTTTTTAAGTTTAGTAACTAATGAGTCGTC"
/label=RS62
112885..112904
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112905..112960)
/note="100% identity to consensus
CAAATTCGTCGGTTTCAGCGGTTTCGATAAATCACTGTAACGTTAAGTTCCCGG"
/label=RS41
112961..112980
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(112981..113031)
/note="99% identity to consensus
GACGGCGTAAAGGTTCTGTGTTTTTAAGTTTAGTAACTAATGAGTCGTC"
/label=RS62
113032..113051
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(113052..113107)
/note="82% identity to consensus
CCAGTCGGTCAGTTTCAGTCATTTCCGATAAATTCCTGTGCTTTTCATTCTAG"
/label=RS17
113108..113127
/note=">= 90% match to ATCCCNNNNNNNNGGGAAT"
/label=DRS3
complement(1131274..113319)
113296..113305
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(113332..115626)
/genes="NMA1161"
/genes="NMA1161"
/note="NMA1161, probable outer-membrane receptor protein,
len: 764 aa; shows weak similarity to many outer-membrane
receptors e.g. TR:O68881 (EMBL:AF055999), phur,
Pseudomonas aeruginosa outer membrane hemin receptor (764
aa), fasta scores; E(): 5.3e-09, 22.5% identity in 746 aa
overlap and TR:P7121 (EMBL:D28119) Pseudomonas aeruginosa
outer membrane protein C precursor (723 aa), fasta scores;
E(): 0.00075, 21.0% identity in 794 aa overlap. Similar to
NMA0474, fasta scores; E(): 2e-05, 20.8% identity in 795
aa overlap. Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative outer-membrane receptor protein"
/protein_id="CAB84423.1"
/db_xref="GI:7379854"
/db_xref="SPTREMBL:O9JUS2"
/translation="MRREAKMAQTTLKPIVLSTILLINTPLLSQAHGTEQSVGLTVSV
VGSRRPRATSLUHTSTADSKITSGDTLROKAVNLGDLGVPGLIHASQYGGASAPV
LQGTGRRIRKVLNHGGETGDMFSPDHAIMVDSALSQQVEILRGPVTLTLYSGNVAG
LVDVADGKIPKMPENGVSGLRLSSGNLEKLTSGGINIGLKNPFLVHTEGLYSGVAG
GDYAPRYRLKRLPDSHADTSOTSGISLWVGKGFIGAAYSDRRDQYGLPAHSHVD
DCHADIIWOKSLINKRYLLOLYPHLLTEEDIDYDNGLSGCFHDDDDAHAAHNGKFWI
DLNRKRYELRAEHWKQFPFGFEALRVHLNRNDYRHKDEKAGDAFFNFQNTQNAIRIELR
HQPIGRILKGSWGVOYLQKSSALSATSEAYKQPMLLDNKVQHSYFSGVQANWDFTL
EGQVRVEKOKASIRYDKALIDRENYNHPLDIGHAHRFALSFGNWFYTPQHLKLS
LTAHQERLPTOELYAHGKHVATNTFEVGNKHLNKRNNIELALGYEGDRQWYLA
LYNRFGNTIYAQTLNDGRPKSIDEDSEMKLVRYNOSGADFYGAGEIYFKPTPRV
IGVSGDYVRGLKNLPSLPGREDAYGNRPLIAODADONAPRPAARLGVHLKASLTORI
DANLDYRYFAONKRLARYETRPFGHMLNLGANYRRNTRYIGENWYTKADNLLNQSVY
AUSFFLSDTTPQMRSPSTGGVNVKF"
complement(113717..113726)
/genes="NMA1161"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(114428..114437)

```

```

misc_feature
/label=DUS
misc_feature
/label=DUS

```

/note="NMA1156, len: 76 aa; unknown, lysine-rich"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1156"
/protein_id="CAB84418.1"
/db_xref="GI:7379849"
/db_xref="SPTREMBL:Q9JUS7"
/translation="MWRSHRGVGVAGYGLLQKLLKKPKPTPLKRLKKKQKQTKRMP
KLLIKTKKKNNSDSETEKRIPIMGNPFVL"
complement(106211..106220)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
106231..106240
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(106340..107137)
/gene="NMA1157"
complement(106340..107137)
/gene="NMA1157"
/note="NMA1157, len: 265 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1157"
/protein_id="CAB84419.1"
/db_xref="GI:7379850"
/db_xref="SPTREMBL:Q9JUS6"
/translation="MQAFOISPAVATRFALYCSSFSKNLRFDTKSKFYLSTEDPEPQ
FPHSEATKILNGSLIFDIDSOHQLPFPDLPVLIINPDVLIFSKSALKSLKPLK
ISYHISSEKSVIPLALVNDGFWRHLRAFRDKNRFADFLVTRIIQANILLAEIN
ENETINNDIDWNRIVKIELVPHPNHDEAKVIEMDFGMTGKLELALRASNIGYVLQ
QNRVDCSDHHIDDPAFRLWLLDNLVLYGINNAEHLAPAFDKKRSKP"
complement(107861..108049)
/gene="NMA1158"
complement(107861..108049)
/gene="NMA1158"
/note="NMA1158, len: 62 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1158"
/protein_id="CAB84420.1"
/db_xref="GI:7379851"
/db_xref="SPTREMBL:Q9JUS5"
/translation="WALKSLKTLNPKQVNIKGIPHFISEITLNLISNNVRMPDKAIK
IPKNKTLIIHWSYLEFF"
108072..108108
complement(108106..110952)
/gene="uvrA"
complement(108106..110952)
/gene="uvrA"
/note="NMA1159, uvrA, excinuclease ABC subunit A, len: 948
aa; similar to many e.g. SW:UVRA_NEIGO (EMBL:U34760),
uvrA, Neisseria gonorrhoeae excinuclease ABC subunit A
(950 aa), fasta scores: E(): 0, 97.4% identity in 948 aa
overlap. Contains two Pfam matches to entry PF00005
ABC_tran, ABC transporter, two PS00017 ATP/GTP-binding
site motif A (P-loop) and two PS00211 ABC transporters
family signature"
/codon_start=1
/transl_table=11
/product="putative excinuclease ABC subunit A"
/protein_id="CAB84421.1"
/db_xref="GI:7379852"
/db_xref="SPTREMBL:Q9JUS4"
/translation="MCNHHPKSHGNDTIRIGARTHNLKNVDLDIPRHKLVVVTGLS
GSGKSLAFDTLYAEGORYVESLSAYAROFLOMDKPDVDLIEGLSPAISLEQKSTS
HNPRSTGVTEIHDYLLRYARVCTPYCPPEHKLPSSOTVSQMDAVLKLPEDTRVM
ILAPAVRRKGEVDFFADLQAGQFARVRVDGEVYOLDEVPKLEKNKHNDIVIDRV
KYKDIKRLAESFTALRHGNEALAMEMDSGEHWFSAFACPVCSYSLPELPRL
FYNPNMSECTDGLGNTNFPDPEKVVVAHPELSAIGDWRKNQFYFOMIQSLA
RHYGFDVQAAMETLPAKIKVVLHGSGKEV IDFTYLSERGTTFNRSFAFEGIIIPLER
RYRETDSETVREKLEQYQNHRCAPSCGGARLKEARYYVVSCEPLHEYSAWFLTKTHR

FFETLDGNKKQIAEKILKEITERLGLFINVGLDYLNLRSAAETLSGGEAQRILAS
QIGSGLTGVMYLDPEISGLHORDNRLIATLKLRLDGLNSVIVVSHDDDAIREADFV
VDMGPCAGEHGVNLIADTPENVAKCENSVTQYLSGKSIAPVSPRTVPNCRMLVL
KCARGNNLKNVLEPLGLGLTICITGVSGSGKSTLINDTLAKITARELNRAQEPAPYD
DIRGLEHLDKVINVDQSPIGRTPRSPNATYVGTPTIRELFASVPLSRERGVNVRFS
FNVKGRCACQGDGVIKVEMHFLPDVYVPCEVCHGRKNYRRETLEIYQKGNISQVLD
MYVEACEFEDAVPTVSRKLOTLMDVLGYIRLGQSATTLSGGEAQRVLALSLSKRD
TGRTLYILDEPTTGLHFADIALLELVIGRLKKGKNSIVITIEHNLDVIKTADWIVDLGP
EGDGGGRIIASGSPPEEVAKVGSYTGKYLKNLL"
complement(108214..109029)
/gene="uvrA"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 122.90, E-value 5.9e-33"
complement(108293..108302)
/gene="uvrA"
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(108397..108441)
/gene="uvrA"
/note="PS00211 ABC transporters family signature"
complement(108625..108642)
/gene="uvrA"
/note="PS00190 Cytochrome c family heme-binding site
signature"
complement(108985..109008)
/gene="uvrA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(109117..109126)
/gene="uvrA"
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(109237..109899)
/gene="uvrA"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 7.70, E-value 6e-06"
complement(109423..109467)
/gene="uvrA"
/note="PS00211 ABC transporters family signature"
complement(110806..110829)
/gene="uvrA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
111003..111038
/note="Stem loop containing DNA uptake sequences: cggat
gccgtctgaa gccgcg ttcagacggc atttg"
111008..111017
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(111024..111033)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
111097..111100
111107..111904
/gene="NMA1160"
111107..111904
/gene="NMA1160"
/note="NMA1160, possible membrane protein, len: 265 aa;
similar to bacterial hypothetical proteins e.g. TR:Q50967
(EMBL:U34760) Neisseria gonorrhoeae hypothetical protein
(259 aa), fasta scores; E(): 0, 93.2% identity in 265 aa
overlap. Contains a hydrophobic, possible membrane anchor
at the N-terminus"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84422.1"
/db_xref="GI:7379853"
/db_xref="SPTREMBL:Q9JUS3"
/translation="MNRLYPHPIIARECHPIIGGGLALSLLVSMCCGWSLFPWVFTV
FALQFFRPAREIPONPEAVLSVDGRIVVVARPDYRDVDAIKISIFMNVFNHSQ
KSPADCTTKVYVNGKFVNADLDKASTENRANVLAFTASGRETTFVYQVAGLVARRI
LCYTQAGTKLSRGERYGFIIRFGSRVMDYLPVDAQVQAIQDKGVNYSVTIARLPLTAP
QIESPESPALQAPVETAAFPSAEQRIEAAVAAKIQAAQVDVLKD"
complement(112016..112071)
repeat_unit

```

/product="putative dihydrolipoamide dehydrogenase E3
component"
/protein_id="CAB84413.1"
/db_xref="GI:7379844"
/db_xref="SPTREMBL:Q9JUT1"
/translation="MSQYDVVVGAGGGVYVAAIRAAQLGFKTACVDAGVKNKAGNAPA
LGGTLNVCIPSLKALQSSSEHFAAEHAEHGTIVGDVKFDVAKMIEKDAIVTKL
TGGVKFLFQNNKVTSLFGTAFAGKNGDAYQIEVDNRGEKTVIEAKHVIIVATGVSPPR
LPQVADINVNLDNEGALNTEVPKLGIVGSGVIGLEMGVSNRNGVAGVETILEAAPT
FLAADOQIAKEAKFYTKEOGLSIELGVKIGDJKSEKGVSAVETAAAGEAKTEVED
KLIIVAGRIPTNKGNAFAVGLKEDRGFKIVDGEKRTNLPNWAIGDVVRGPMIAHK
ASDEGVAERIAQKPHIDFNVPFVITYDPEIAWVGKTEEQKAKGVYKKTSGF
GANGRALAMGRKGTGVKVLADAKTRDRLGLVHMIGPVVSELLTEGVTALEFFASSEDIA
RIHAHPTLSEVVEHAALAADKRALHG"
/misc_feature
101135..102550
/gene="lpda3"
/notes="Pfam match to entry PF00070 pyr_redox, Pyridine
nucleotide-disulphide oxidoreductase class-I, score
680.90, E-value 6.3e-201"
/misc_feature
101255..101287
/gene="lpda3"
/notes="PS00076 Pyridine nucleotide-disulphide
oxidoreductases class-I active site"
102648..102935
/gene="NMA1152"
102648..102935
/gene="NMA1152"
/notes="NMA1152, possible membrane protein, len: 95 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions and a possible N-terminal signal sequence"
/misc_feature
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84414.1"
/db_xref="GI:7379845"
/db_xref="SPTREMBL:Q9JR20"
/translation="MKNEIVGIFFPAGIISMCAALWQMYVMVTYITLNRFKREL
VWRVALLTFSLSAVYLLCPNRRKKGIVFPIILGGGAAYLLARMWLPFSK"
103023..103027
103034..104200
/gene="sucC"
103034..104200
/gene="sucC"
/EC_number="6.2.1.5"
/notes="NMA1153, sucC, probable succinyl-CoA synthetase
beta subunit, len: 388 aa; similar to many e.g.
SW:SUCC_ECOLI (EMBL:J01619), succ, Escherichia coli
succinyl-CoA synthetase beta chain (EC 6.2.1.5) (388 aa),
fasta scores; E(): 0, 60.5% identity in 387 aa overlap.
Contains Pfam match to entry PF00549 ligase-CoA,
CoA-ligases and PS01217 ATP-citrate lyase / succinyl-CoA
ligases family signature 3"
/codon_start=1
/transl_table=11
/product="putative succinyl-CoA synthetase beta subunit"
/protein_id="CAB84415.1"
/db_xref="GI:7379846"
/db_xref="SPTREMBL:Q9JUT0"
/translation="MNLHEYQAKELASGLPVOGGILAHNGEERAAAYDKLGGKFAV
VKQVHAGRGKAGGVKVKSREREKVAESLIGTLNLTVTQDANGOPVNSVLVCEDM
VPQTEILGAVVDVDRTRRTFMASGTGGVEIEKVAETPEKIFKVTPLVGLQPCQ
ARVAFOLGDKQINFEVFKLMTGAYKAEVNDFALEFVNPLAVRENGALACVDGKIG
IDSNALYRLPKLAELRDKSQENERELKSEFDLNVVALEGNICGMVNGACMATWDI
IKLKGQPAFLDVGGAATKDRVVEAFKLILEDKSVKGLVINIFEGIVRCGLMATAEIV
AAVKELNDVPVVRLEGNNAELGAKILNESGLKLTSADGLNDAEKIIVAAVNA"
103772..104191
/gene="sucC"
/notes="Pfam match to entry PF00549 ligase-CoA,
CoA-ligases, score 259.70, E-value 3.8e-74"
103805..103879
/gene="sucC"
/notes="PS01217 ATP-citrate lyase / succinyl-CoA ligases
family signature 3"

```

```

RBS
gene
104200..104204
104211..105101
/gene="sucD"
104211..105101
/gene="sucD"
/EC_number="6.2.1.5"
/notes="NMA1154, sucD, probable probable succinyl-CoA
synthetase alpha subunit, len: 296 aa; highly similar to
many e.g. SW:SUCD_ECOLI (EMBL:J01619), sucD, Escherichia
coli succinyl-CoA synthetase alpha chain (EC 6.2.1.5) (288
aa), fasta scores; E(): 0, 72.1% identity in 290 aa
overlap. Contains Pfam match to entry PF00549 ligase-CoA,
CoA-ligases, PS01216 ATP-citrate lyase / succinyl-CoA
ligases family signature 1 and PS00399 ATP-citrate lyase /
succinyl-CoA ligases family active site"
/codon_start=1
/transl_table=11
/product="putative succinyl-CoA synthetase alpha subunit"
/protein_id="CAB84416.1"
/db_xref="GI:7379847"
/db_xref="SPTREMBL:Q9JUS9"
/translation="MSVLINKTKVLVQGTGKNGTFHSEQALAYGTVKVGVPGRKG
GQTHLDLPVNTKAEKVCADASVIYPAPFVLDSIVAEVDSGLVWVITGVPT
LDMLKAKRYLETNGTRLVGPCPGVITPECKIGIMPGRHITPGRIGIISRGTLT
YEAAOTTKILGLOGSTCIGIGDPIPCMNQIDALKLQEDPDPTDAIMIGEIGGTARE
EAAEYIQSNVSKPVGYIAGVTAPKGRMGHAGAILISGGRTAEKFAEKAIAIYT
RSPAELGTTMLEVLTKTKGLA"
104613..105056
/gene="sucD"
/notes="Pfam match to entry PF00549 ligase-CoA,
CoA-ligases, score 210.80, E-value 2.1e-59"
104670..104759
/gene="sucD"
/notes="PS01216 ATP-citrate lyase / succinyl-CoA ligases
family signature 1"
104922..104963
/gene="sucD"
/notes="PS00399 ATP-citrate lyase / succinyl-CoA ligases
family active site"
105127..105170
/notes="Stem loop containing DNA uptake sequences: caaat
gcgcgtctgaa accggaatacgggt ttcacagcgc atttt"
105132..105141
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(105156..105165)
/label=DUS
/notes="Core DNA uptake sequence: gcgcgtctgaa"
complement(105336..105698)
/gene="NMA1155"
complement(105336..105698)
/gene="NMA1155"
/notes="NMA1155, len: 120 aa; unknown, shows similarity to
parts of hypothetical proteins e.g. SW:YAFM_HAEIN
(EMBL:U32706), HI0217, Haemophilus influenzae hypothetical
protein (176 aa), fasta scores; E(): 2.6e-15, 42.5%
identity in 106 aa overlap. Similar to parts of NMA1675,
fasta scores; E(): 8.4e-07, 42.7% identity in 103 aa
overlap and NMA0532, fasta scores; E(): 8.4e-07, 42.7%
identity in 103 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1155"
/protein_id="CAB84417.1"
/db_xref="GI:7379848"
/db_xref="SPTREMBL:Q9JUS8"
/translation="MSNISASRAAYMDVOKYPFETVAVCVLPNHIIHTWLPDDAD
YSLIRLIIKTSAYSPTKKNLSAGKQHERGIRRFYGHYTRDETDLQRCADTIH
SNPIKQMSRAKTESCKK"
105944..106174
/gene="NMA1156"
105944..106174
/gene="NMA1156"

```

AIYRLISKIPTIAAMCYRYSNGLPFNYPKNNLNLSYSENFLHMMFATPCEDYKPNPVLAR
ALDRIFILHADHEONASTSTVRLAGSSGANPRACIAAGIACIAGWASHGANGAENVALKML
DEIGDVSNVAYMEGVKORKYRLMCGFHRVYRNMOPRASIMRETCTCYEVLKELGLEDSP
KFKLAMELQIALKDPFFIERKLYPNVDYISGYISALSIGPTMETFIVIFALSRVGVWI
SWHEMISDPSLKIQRPROLYTGSERRDYVPPGER"
95263. .96354
/gene="gita"
/note="Pfam match to entry PF00285 citrate_synth, Citrate
synthase, score 730.40, E-value 2.3e-229"
96028. .96066
/gene="gita"
/note="PS00480 Citrate synthase signature"
96616. .99444
/gene="sucA"
96616. .99444
/gene="sucA"
/EC_number="1.2.4.2"
/note="NMA1149, sucA, probable 2-oxoglutarate
dehydrogenase E1 component, len: 942 aa; similar to many
E1 components of 2-oxoglutarate dehydrogenase complex e.g.
SW:ODO1_ALCEU (EMBL:X91877), odhA, Alcaligenes eutrophus
oxoglutarate dehydrogenase (EC 1.2.4.2) (950 aa), fasta
scores: E(): 0, 54.5% identity in 945 aa overlap. Contains
Pfam match to entry PF00676 E1_dehydrog, Dehydrogenase E1
component"
/codon_start=1
/transl_table=11
/product="putative 2-oxoglutarate dehydrogenase E1
component"
/protein_id="CAB84411.1"
/db_xref="GI:7379842"
/db_xref="SPTREMBL:O9JRJ8"
/translation="WMDEKLNESYLFGSNAPVIELELYEAFLENPDVADDEKWKQYETDL
VQAGTVADVAVTPIRESFVTLAKKRIASAVAGGADEAMLRKKQVSLRLISAYRIQG
SKQADLPVLRIPPRDIEALDPKFLGLSDAMALQFMNGEGDFANGKLPLSQIISNL
KOTYCGHTALEYIYIPNTEERRWNRNYESVLSLTPHNAQDKRRLIKENTVLAETILRY
LHTKVYQGRKFGVEGESAIAGNLNLIQNAQDGVVEEVIIGMAHGRNLNVLNILGKK
PGDLFAERFGRAEIKLPQSDVKVYHMGFSSDIATPHGPMHVSLAENPSHLEIVNPVVEG
SARAKOKRGENGRDVKLPVLJHGDSAFIGLVGNQATENLSKTRGTYTGGTVHIVNN
QIGFTSIDRTPRSVHCIDIAKMVSAPVIHVNQDGPVRCFAIQALDYRKKFKHDI
VIDVYRKWGHNEGDDPTLTQPMYKVKVSHQPGARALYEQLIAEGVVYQAEADGYI
QAYRALDGRHEVETLTLSNFQRTIDMSKYQCKDMREHIEITGLPAADIERLTERFTA
VPEGALHPTRAKRVIEARKAMASGKQAIIDWGAETLAVASLTKGHRVIRIGEDSGRG
TFSHRHVLHDQREKWDGTVPLRHMGEGMGEFLVIDSLNEAEVMAFEGFACSA
PKLTIWEAQFGDAQOVTIDOLFSSGETKWRGLCGITTLPHCYDQSGQPEHSSAR
VERWLOLSENNKQVIMPEASQMFHLLQROVLGSYRKLPLVIFPMKSKRLRLFRFGANSP
ENFTGSTRPVIGDTAERASNDVKRVVICAGQYVYDLEAGREKLEDDVAIVRVE
QLYPPFYDEVKAEALAKYPNAKSVVWQAQEPKNOGAFYQIRHRIEDVISEEQKLSYAGR
PSSASPAVGYSKHHIAQLKQLEDALAL"
97234. .98214
/gene="sucA"
/note="Pfam match to entry PF00676 E1_dehydrog,
dehydrogenase E1 component, score 120.90, E-value 2.3e-32"
97406. .97415
/gene="sucA"
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
99514. .99523
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
99533. .99537
99544. .100755
/gene="sucB"
99544. .100755
/gene="sucB"
/EC_number="2.3.1.61"
/note="NMA1150, sucB, probable dihydroliipoamide
succinyltransferase E2 component, len: 403 aa; similar to
many E2 components of 2-oxoglutarate dehydrogenase complex
e.g. SW:ODO2_ECOLI (EMBL:J01619), sucB, Escherichia coli
dihydroliipoamide succinyltransferase component of
2-oxoglutarate dehydrogenase complex (EC 2.3.1.61) (404
aa), fasta scores: E(): 0, 57.1% identity in 408 aa

overlap. Similar to NMA1555, fasta scores: E(): 8.9e-24,
35.8% identity in 433 aa overlap. Contains 3x degenerate
TAAAPAAA repeats at aa 79-113. Contains Pfam matches to
entry PF00198 2-oxoacid_dh, 2-oxo acid dehydrogenases
acyltransferase (catalytic domain) and to entry PF00364
biotin_lipoyl, Biotin-requiring enzymes and PS00189 2-oxo
acid dehydrogenases acyltransferase component lipoyl
binding site"
/codon_start=1
/transl_table=11
/product="putative dihydroliipoamide succinyltransferase E2
component"
/protein_id="CAB84412.1"
/db_xref="GI:7379843"
/db_xref="SPTREMBL:O9JUT2"
/translation="MIIDVKVPMLSSEVSEGTLLLEWKVKVGEVARDEILIDIETDKV
VLEVPSPQAGVLVEIVAOGGETVADQVLARIIDTAATAAAEAPAAATAAEPAAAA
EAAAPAAAPAAONNAAMPAAAKIAESGVVDNALQSGRDGRVLEKDVQNAAKPAAA
AAPAVALPAGARPEERVPMSRLRARVAERLLASQENAILTTFNEVNMKPIMDLRAKY
KKKFEHGVKLGSEFFVKAAVAALKYPVVNASVDGKDIVVHGVFDIGIAGSPRG
LVVPIILRDADOMSIADIEQALIVDYAKKADGKIATIEDLTGTFSTFINGCTFGSMSTP
IINPQSAILGHMHTKRAVWENGQVVVRPMMYLSLYSDHRIIDGREAVLTILVAIKDA
LEDPARLLLDL"
99550. .99771
/gene="sucB"
/note="Pfam match to entry PF00364 biotin_lipoyl,
Biotin-requiring enzymes, score 102.80, E-value 6.6e-27"
99622. .99711
/gene="sucB"
/note="PS00189 2-oxo acid dehydrogenases acyltransferase
component lipoyl binding site"
100090. .100749
/gene="sucB"
/note="Pfam match to entry PF00198 2-oxoacid_dh, 2-oxo
acid dehydrogenases acyltransferase (catalytic domain),
score 420.10, E-value 2e-122"
complement(100760. .100769)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
100791. .100800
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
complement(100805. .100847)
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 50.38"
/label=Correia
complement(100849. .100911)
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 78.71"
/label=Correia
complement(101011. .101020)
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
101109. .101112
101120. .102553
/gene="lpdA3"
101120. .102553
/gene="lpdA3"
/EC_number="1.8.1.4"
/note="NMA1151, lpdA3, probable dihydroliipoamide
dehydrogenase E3 component, len: 477 aa; similar to many
E3 components of 2-oxoglutarate dehydrogenase complex e.g.
SW:DLDH_ALCEU (EMBL:X91877), odhL, Alcaligenes eutrophus
dihydroliipoamide dehydrogenase (EC 1.8.1.4) (474 aa),
fasta scores: E(): 0, 62.3% identity in 472 aa overlap.
Similar to NMA1556, fasta scores: E(): 0, 38.2% identity
in 476 aa overlap and NMA1142, fasta scores: E(): 5.5e-26,
28.9% identity in 481 aa overlap. Contains Pfam match to
entry PF00070 pyr_redox, Pyridine nucleotide-disulphide
oxidoreductase class-I and PS00076 Pyridine
nucleotide-disulphide oxidoreductases class-I active site"
/codon_start=1
/transl_table=11

/note="NMA1144, sdhD, probable succinate dehydrogenase hydrophobic membrane anchor protein, len: 113 aa; similar to SW:DHSD_ECOLI (EMBL:J01619), sdhD, Escherichia coli succinate dehydrogenase hydrophobic membrane anchor protein (115 aa), fasta scores; E(): 5e-13, 36.5% identity in 104 aa overlap. Also similar to TR:O83000 (EMBL:AB015757) Rhodoferrax fermentans fumarate reductase cytochrome B small subunit (121 aa), fasta scores; E(): 1.8e-21, 48.6% identity in 111 aa overlap. Contains hydrophobic, probable membrane-spanning regions" /codon_start=1 /transl_table=11 /product="putative succinate dehydrogenase hydrophobic membrane anchor protein" /protein_id="CAB84406.1" /db_xref="GI:7379837" /db_xref="SPTREMBL:Q9JUT4" /translation="MVERKLTGAHYGLRDWAMQRTAVIMLIYVALIVLALPKREY SAWQAFSTQWVKVFTQVSFIATVFLHAWVGIRDLWMDYIKPFGVRLQLQVATIVLWVG CLVYSIKVIWG" 92174. .93937 /gene="sdhA" /EC_number="1.3.99.1" /note="NMA1145, sdhA, probable succinate dehydrogenase flavoprotein subunit, len: 587 aa; similar to many e.g. SW:DHSA_ECOLI (EMBL:J01619), sdhA, Escherichia coli succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1) (588 aa), fasta scores; E(): 0, 56.6% identity in 590 aa overlap. Also similar to TR:O83001 (EMBL:AB015757) Rhodoferrax fermentans fumarate reductase flavoprotein subunit (601 aa), fasta scores; E(): 0, 64.2% identity in 592 aa overlap. Similar to NMA2092, fasta scores; E(): 8.6e-31, 33.9% identity in 546 aa overlap. Contains Pfam match to entry PF00890 FAD_binding_2, FAD binding domain and PS00504 Fumarate reductase / succinate dehydrogenase FAD-binding site" /codon_start=1 /transl_table=11 /product="putative succinate dehydrogenase flavoprotein subunit" /protein_id="CAB84407.1" /db_xref="GI:7379838" /db_xref="SPTREMBL:Q9JUT3" /translation="MGFPVKKFDVAVIVGGGAGLRAALQLSKLNCVLSKVFPTRS HTVAAGQGSASLGNVOEDRWMDYDTVKGSDWLQDDAIEFMCRAAPPAVIELEHM GMPFVRESGIYQRPFGGHTAEHGKRAVERACAVADRTGHAMLTLYQONVRANTQF FVEMTAQDLIRNDENGVDVGTAMEMETGEVY IFHAKAVFATGGGRIVASSTNAYMN TGDGLGICARAGIPIEDMEFQHPHTGVAGAGVLITEGVREGGILLNADGERFMERY APTVKLASRDVVSRAAMEIYEGRCGNKQKHVLLKIDHIGAEKIMEKLPGIREISI OFAGIDPIDPIPVPTTHYMMGGIPTNVHGEVYVFGDEYEPVKGLXAAGECACAS VHGANRLGNSLLDLVYFGKAAGDSMKIFKEQSDWKPLPANAGELTRQRIERLDNOT DGNVDALRELQRSVOLHAGVFTDEILSKGVREVMATAEVRKTEIKDKSKVWNTA RIEALDELNIEVAKATLVSAEARKESRGAAHSDDDHPDERDENMKHTLYHSANTLS YKPVHTKPLSVEYIKPAKRVY" 92300. .92329 /gene="sdhA" /note="PS00504 Fumarate reductase / succinate dehydrogenase FAD-binding site" 92495. .93490 /gene="sdhA" /note="Pfam match to entry PF00890 FAD_binding_2, FAD binding domain, score 664.90, E-value 4.2e-196" 94044. .94047 /gene="sdhB" 94057. .94764 /gene="sdhB" 94057. .94764 /EC_number="1.3.99.1" /note="NMA1146, sdhB, probable succinate dehydrogenase iron-sulphur protein, len: 235 aa; similar to many e.g. SW:DHBS_ECOLI (EMBL:J01619), sdhB, Escherichia coli

misc_feature

misc_feature

RBS

gene

CDS

succinate dehydrogenase iron-sulphur protein (EC 1.3.99.1) (238 aa), fasta scores; E(): 0, 66.0% identity in 235 aa overlap. Also similar to TR:O83002 (EMBL:AB015757) Rhodoferrax fermentans fumarate reductase iron-sulphur protein subunit (234 aa), fasta scores; E(): 0, 72.8% identity in 235 aa overlap. Contains Pfam match to entry PF00037 fer4, 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains and PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature" /codon_start=1 /transl_table=11 /product="putative succinate dehydrogenase iron-sulphur protein" /protein_id="CAB84408.1" /db_xref="GI:7379839" /db_xref="SPTREMBL:Q9JRI0" /translation="MEKMSFEIYRNPDDVDAKPYMORYELEPTDYKLLDALVRLKA ODDTLSFRRCRSGICGSDGMNINGLACLTDLRGLKOPVKIRPLGLPVIKRLDILV DMTQFFKQYHSVKPYVYNDNPIDADKRGLOTOERKELDGLYECILCACSTACPSFW WNPDKFVCPGSGLLNAYRFIADSRDITITNERLDNLNDPYLFRCHTIMNCVDVCPKHLN PTRAIGKIKIEMIKRAV" 94471. .94719 /gene="sdhB" /note="Pfam match to entry PF00037 fer4, 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains., score 53.10, E-value 3e-14" 94492. .94527 /gene="sdhB" /note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature" 94768. .95016 /gene="NMA1147" 94768. .95016 /gene="NMA1147" /note="NMA1147, len: 82 aa; shows weak similarity to bacterial hypothetical proteins e.g. TR:Q92CV9 (EMBL:AJ235272), RPS99, Rickettsia prowazekii hypothetical protein (87 aa), fasta scores; E(): 0.00085, 30.7% identity in 75 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1147" /protein_id="CAB84409.1" /db_xref="GI:7379840" /db_xref="SPTREMBL:Q9JR91" /translation="MVVFDIAKRRKIRFQTRRGLLELDLIFGRFMEKEFEHLSDKELS EFSEILFQDQELLALINGHSETDKGLHLPMEKIRRA" 95059. .95068 /note="Core DNA uptake sequence: gcgcgtctgaa" /label=DUS 95116. .95120 /gene="glta" 95128. .96411 /gene="glta" 95128. .96411 /gene="glta" /EC_number="4.1.3.7" /note="NMA1148, gltA, probable citrate synthase, len: 427 aa; similar to many e.g. SW:CISY_PSEAE (EMBL:M29728), gltA, Pseudomonas aeruginosa citrate synthase (EC 4.1.3.7) (428 aa), fasta scores; E(): 0, 69.6% identity in 427 aa overlap. Similar to NMA2054, fasta scores; E(): 1.4e-31, 31.8% identity in 362 aa overlap. Contains Pfam match to entry PF00285 citrate_synth, Citrate synthase and PS00480 Citrate synthase signature" /codon_start=1 /transl_table=11 /product="putative citrate synthase" /protein_id="CAB84410.1" /db_xref="GI:7379841" /db_xref="SPTREMBL:Q9JQX0" /translation="MSKSIKLVNPGGAGLEPLVLEASIGHDVVDIRGLTKNTGLFSFD PCGFVTSASCEKITYIDGOGLLYYRGYPIEQIAEKSDYLEVCYLLIYELGPTPEKKA EFDNTVRRHTWHEQTLTWFFRFRRAHPMMVMGVVVGALSAFYQDSLDISNPEHRKI

/gene="metE"
86236. .88512
/gene="metE"
/EC_number="2.1.1.14"
/note="NMA1140, metE, probable
5-methyltetrahydropteroyltriglutamate--homocysteine
methyltransferase, len: 758 aa; similar to many e.g.
SW:MEYE_ECOLI (EMBL:M87049), metE, Escherichia coli
5-methyltetrahydropteroyltriglutamate--homocysteine
methyltransferase (EC 2.1.1.14) (752 aa), fasta scores:
E(): 0, 55.1% identity in 757 aa overlap"
/codon_start=1
/transl_table=11
/product="putative
5-methyltetrahydropteroyltriglutamate--homocysteine
methyltransferase"
/protein_id="CAB84402.1"
/db_xref="GI:7379833"
/db_xref="SPTREMBL:Q9JUT6"
/translation="MTLHFGPPRVGAFRELFAOEKYWRKETISQELLAVAKDLRE
KNWKHOAANADYVAVGDETFYDHI DLQVATGAI FARGFDSQNLSLQFFQLARGN
KQGFLEMTKWFDTNTHILVPEPHADTEFKANAKHYVQQLEQAQLGLKAKPTVVGPL
TFLWQKEKGAVEDRLSLPLKLLPVYVELTALVEAGAEMIQIDEPALTVDLPKEWY
EAYKDYATLSKYSAKILLSTYFGSVAEHAALLKSLPVDGLHIDLVRAPQLDAFADY
DKVLSAGVTDGRNIWRANLKVLETVELLAQKLGDRLWISSCSLLHTPDLSEVKL
KANKPLYSLAFTLOKTOELRVLKAALNEGRDSVAEELAAASOAAADSRRANSEIHRK
DVAKRLADLPANADQKSPFADRIKAAQWNLPLPNTNIGSFEPQTETIRQARAFK
KGLSRLADYEAMKKEIALVFEQEKLDLDLVHGEARNDMVEYFCELLSGFAFTQY
GWQDYSRCVKRPPIIFGDVSRPEAMTVAMSTYAQSLTRPKMKGLTGPVTILQWSEV
RNDIPRSTGCKIOTALNDEVLDELRKAGIKVQIDEPAIRREGPLKRADWDLYNNAG
ESPRLSGCEDSTQIHTMCSYFENDILPAIAAMDADVITITSRSDMELLTAFGEF
KYPNDIPGPGYDIHSPRVPTAEVHELRLKAIIEVVPVRLWVNPDCGLTRGWKETLE
QLQVMNVTHKRLAEALAK"
88529. .88563
/note="Stem loop containing DNA uptake sequences: taaat
accgtctgaa agcct ttacagcgt atttt"
RBS
88780. .88784
gene
88791. .89528
/gene="NMA1141"
88791. .89528
/gene="NMA1141"
/note="NMA1141, possible redoxin, len: 245 aa; similar to
many from both bacteria and eukaryotes e.g. TR:AAD28242
(EMBL:AF121355) Arabidopsis thaliana peroxiredoxin (162
aa), fasta scores: E(): 2e-17, 39.9% identity in 158 aa
overlap. Contains Pfam match to entry PF00578 AhpC-TSA,
AhpC/TSA family"
/codon_start=1
/transl_table=11
/product="putative redoxin"
/protein_id="CAB84403.1"
/db_xref="GI:7379834"
/db_xref="SPTREMBL:Q9JQS4"
/translation="MALQDRTGQKVPSSVVFRTRVGDTKWDVSTDLLFKGKKVVFSLP
GAFTPTCCSSHLPRYNELFGAFKENGVDIYCVNSDTEFVMAWAAEESDNIYMPID
GNCEFTGSGMLVKGEDLGFGRKSRYSMLVNDGVVVEKMFIRPEPGDPFKYSDADTM
LQFVAPDKAQSVAIFTRPGQCFCAKAKAQALQDKLSVEEIVLGKDAVTVSRAITG
KTAPOVFGIGYVIGGSEBLEAYLARN"
88809. .89276
/gene="NMA1141"
/note="Pfam match to entry PF00578 AhpC-TSA, AhpC/TSA
family, score 157.90, E-value 1.7e-43"
89713. .89722
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
89765. .89768
gene
89777. .91180
/gene="lpdA2"
89777. .91180
/gene="lpdA2"
/EC_number="1.8.1.4"
/note="NMA1142, lpdA2, probable dihydrolipoamide
dehydrogenase, len: 467 aa; similar to many e.g.

SW:DLDL_BACST (EMBL:X53560), pdhD, Bacillus
stearothermophilus dihydrolipoamide dehydrogenase (EC
1.8.1.4) (470 aa), fasta scores: E(): 0, 30.0% identity in
454 aa overlap. Similar to NMA1151, fasta scores: E():
3.5e-28, 28.6% identity in 482 aa overlap and NMA1556,
fasta scores: E(): 1.9e-24, 26.0% identity in 469 aa
overlap. Contains Pfam match to entry PF00070 pyr_redox,
pyridine nucleotide-disulphide oxidoreductase class-I."
/codon_start=1
/transl_table=11
/product="putative dihydrolipoamide dehydrogenase"
/protein_id="CAB84404.1"
/db_xref="GI:7379835"
/db_xref="SPTREMBL:Q9JUT5"
/translation="MKKIQADVIVGGTAGMCAFRNARLHSDNYLIENNVEGTTCFA
RVGCMPSKLLIAAAEARHHAHTDPFGVHLDKDSIVVNGEEVMQRVKSEDRFVGFFV
ADVEEWPADKRIMGSAKFIDEHTVQIDETQITAKSFVIATGSRPILPQWQSLGNRL
IINDVPFSDWTLPKRVAVFGPGVIGLEQLGQALHRLGVKEIFGLGGTIGGIDPVPVSD
EAKAVFGEELKHLDAKTEAKLDADGNVEVHPQDGEKGVFAEYMLAAVGRPRNVSD
IGLENTINIEKDARGVPVADPLTMQTSIPHFIAGDASNLPLLHEAADCKIKIADNAG
RPNIGSGLRRSTIGVVFVTSPOIGFVGLKYAQVAQYQADEFVIGEVSEFNQGRSRVM
LVNKGHMRLYAEKATGRFIGAIEIVGPAAEHLAHLAWAHQMKMTVPQMLDMPFTHPIV
EEGLRTALRDADAKLKA"
89798. .91177
/gene="lpdA2"
/note="Pfam match to entry PF00070 pyr_redox, Pyridine
nucleotide-disulphide oxidoreductase class-I, score
111.00, E-value 2.3e-29"
91194. .91228
/note="Stem loop containing DNA uptake sequences: acaat
gccgtctgaa atttt ttacagcgc atttt"
91199. .91208
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(91214. .91223)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
91459. .91836
/gene="sdhC"
91459. .91836
/gene="sdhC"
/note="NMA1143, sdhC, probable succinate dehydrogenase
cytochrome B subunit, len: 125 aa; similar to many e.g.
SW:DHSC_ECOLI (EMBL:J01619), sdhC, Escherichia coli
succinate dehydrogenase cytochrome B-556 subunit (129 aa),
fasta scores: E(): 1.4e-14, 39.8% identity in 123 aa
overlap. Contains Pfam match to entry PF01127 Sdh_cyt,
Succinate dehydrogenase cytochrome b subunit and PS01000
Succinate dehydrogenase cytochrome b subunit signature 1.
Contains hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative succinate dehydrogenase cytochrome B
subunit"
/protein_id="CAB84405.1"
/db_xref="GI:7379836"
/db_xref="SPTREMBL:Q9JRJ0"
/translation="MSAKPRPVTLDPNLRPLIPGVISILHRSISGVGLFIMLPFLFLYF
LSGTLSQESAFETYRAIVSHPLVKVLIGVLMAYLHHSLAGIRFLDLAHRKGLELNTA
RNTAKAVFASALVLTVVLGALLW"
91465. .91815
/gene="sdhC"
/note="Pfam match to entry PF01127 Sdh_cyt, Succinate
dehydrogenase cytochrome b subunit, score 79.90, E-value
5.3e-20"
91474. .91548
/gene="sdhC"
/note="PS01000 Succinate dehydrogenase cytochrome b
subunit signature 1"
91830. .92171
/gene="sdhD"
91830. .92171
/gene="sdhD"

family, PS00665 Dihydrodipicolinate synthetase signature 1
and PS00666 Dihydrodipicolinate synthetase signature 2"

/codon_start=1
/transl_table=11
/product="putative dihydrodipicolinate synthase"
/protein_id="CAB84386.1"
/db_xref="GI:7379818"
/db_xref="SPTREMBL:Q9JU09"
/translation="MLGSLVALITPMNODGSIHYEQRLRIDLWHIENGTDGIVAGVT
TGESATISVEHETAVIEAVYKHKRVPVIAGTGANNITVEAIALSQAERAGADYTL
VYPNPNPSQGMNHRFAKAAEAAILPMILYNVFGRTVSMNNTILRLAEIPNIVG
KEASNGNISIELNRAPEGVLSGDDHTALPMLCGGHGIVITVAANAAPKFLADMC
RAALQGDIALARELNDRLIPTIDYTFMFCPSAPAKWAVSALGRCPEHVRPLVPLTEG
GOAKVRAALKASQGL"
complement(73164..73919)
/gene="dapA"
/note="Pfam match to entry PF00701 DHDPs,
Dihydrodipicolinate synthetase family, score 426.60,
E-value 2.3e-124"
74204..74213
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
/gene="NMA1126"
/label="NMA1126"
len: 463 aa; similar to many probable transmembrane transport protein,
eukaryotes e.g. SW:UAPC-EMENI (EMBL:X79796), uapC,
Emicella nidulans purine permease (580 aa), fasta
scores; E(): 4.4e-22, 30.7% identity in 499 aa overlap.
Also similar to hypothetical proteins e.g. SW:YGFO-ECOLI
(EMBL:U28375), ygfO, Escherichia coli hypothetical protein
(485 aa), fasta scores; E(): 0, 69.4% identity in 438 aa
overlap. Contains Pfam match to entry PF00860
xan_ur_permease, xanthine/uracil permeases family and
PS01116 xanthine/uracil permeases family signature.
Contains hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative transmembrane transport protein"
/protein_id="CAB84387.1"
/db_xref="GI:7379819"
/db_xref="SPTREMBL:Q9JO25"
/translation="MSGQLGKGADAPDLVYGLEDRPPFGNALLSAYTHLLAIFVPMIT
PALIVGGAELPVENTAYLSYMAWASGVGYIQVNRFGVPGSMLSIQSNFSEVTV
MIALGAKMKGGLTKDAMISTLLGVSVGAFLVCFSAWLLPYLKVKVITPTVSGVYVL
IGLSLVHVGITDFGGFGAKADGTFGSMENGLASLVLLIVFNCKMKNPLRLMSGTA
VGLIAGYIVALFLGKVDFAQLNQLPLVTLVPVPEKYGFAFDWHAFFIVAGAIFLLSVFEA
VGLDTATAMVSDQPIEGEYTKRLRGVLADGLSVIATAGLSILPTTFAQNGVQIM
TGVASRHVGKYIAVILVLLGLFPVGRFTTIPSPVLGGMVLMFGLIATAGVRLVLS
HGIRREAVTAANTSVGLGLGVAPEPEVFNKLPVLFONISISAGGITAVLLNLVLPEDKT
EAAVKFDTHLEH"
74362..75546
/gene="NMA1126"
/note="Pfam match to entry PF00860 xan_ur_permease,
xanthine/uracil permeases family, score 436.60, E-value
2.1e-127"
75385..75447
/gene="NMA1126"
/note="PS01116 Xanthine/uracil permeases family signature"
75691..75733
/note="Stem loop containing DNA uptake sequences: cgaat
gccgtctgaa acagaatccctgt ttcagaagcc attgt"
75696..75705
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(75719..75728)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(75743..76528)
/gene="NMA1127"

CDS

complement(75743..76528)
/gene="NMA1127"
/note="NMA1127, possible RNA methylase, len: 261 aa; shows
weak similarity to SW:TSNR-STRAZ (EMBL:X02392), tsnR,
Streptomyces azureus RNA
(adenosine-2'-O-)-methyltransferase (EC 2.1.1.66) (269
aa), blastp scores; Expect = 4.0e-14. Similar to other
hypothetical RNA methylases e.g. SW:YJFH-ECOLI
(EMBL:U14003), yjfh, Escherichia coli hypothetical
tRNA/tRNA methyltransferase (243 aa), fasta scores; E():
0, 52.3% identity in 241 aa overlap. Contains Pfam match
to entry PF00588 SpoU_methylase, SpoU RNA methylase
family"
/codon_start=1
/transl_table=11
/product="putative RNA methylase"
/protein_id="CAB84388.1"
/db_xref="GI:7379820"
/db_xref="SPTREMBL:Q9JU08"
/translation="MPSEPFQSQTIMANQRPIYGFHAVNARLWPNKPSIVELYIOEGK
SDARTREVLEKAENETRVYFADADRLNAISKARGHQGVGVGFIDAKNHHVLEVDLEN
LSEPLLILLDGIIDPHNLGACLTADAMGVHIAVPKDKSAGLNATVSKVASGAET
VPYITVNLARTRELKEYGIWIIGTDMGNADLYHCDLPDSAAWVAGNEDGMRRLT
REHCDMLVSIPTMECTVESMVSVSAGVLSLSTRQRVLKNEKA"
complement(75785..76210)
/gene="NMA1127"
/note="Pfam match to entry PF00588 SpoU_methylase, SpoU
RNA methylase family, score 188.10, E-value 1.4e-52"
complement(76517..76526)
/gene="NMA1127"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
76348..76557
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(76562..77050)
/gene="NMA1128"
complement(76562..77050)
/gene="NMA1128"
/note="NMA1128, possible membrane protein, len: 162 aa;
similar to SW:YQGC-BACSU (EMBL:D84432), yqgc, Bacillus
subtilis hypothetical protein (160 aa), fasta scores; E():
3.6e-10, 34.2% identity in 161 aa overlap. Contains
hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84389.1"
/db_xref="GI:7379821"
/db_xref="SPTREMBL:Q9JU07"
/translation="MTALLVILALIAAGTAGIVYPALPGLALMFAGTWLLAYSGGY
QIYGAGVLTWGLISLAGILADYVAGITGKYGAGKLAVRGALAGSIIGIFSLPGL
ILGPFIGAAGELIERRNLQAGKGLTLLGLIVGTAFKIGCAVSILFILLKYIAY
LF"
complement(77055..77501)
/gene="NMA1129"
complement(77055..77501)
/gene="NMA1129"
/EC_number="3.5.4.1"
/note="NMA1129, possible cytosine deaminase, len: 239 aa;
shows weak similarity to many e.g. SW:FCAL-CANAL
(EMBL:U55194), fcal, Candida albicans cytosine deaminase
(EC 3.5.4.1) (150 aa), fasta scores; E(): 5.5e-08, 31.4%
identity in 140 aa overlap. Contains Pfam match to entry
PF00383 dCMP_cyt deam, Cytidine and deoxycytidylate
deaminase zinc-binding region and PS00903 Cytidine and
deoxycytidylate deaminases zinc-binding region signature"
/codon_start=1
/transl_table=11
/product="putative cytosine deaminase"
/protein_id="CAB84390.1"
/db_xref="GI:7379822"
/db_xref="SPTREMBL:Q9JU06"

misc_feature

misc_feature

misc_feature

gene

CDS

gene

CDS

| | | | |
|--------------|--|--------------------|--|
| misc_feature | 69313.. .69861 /gene="NNA1120" /note="Pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 191.70, E-value 1.2e-53" 69313.. .70134 /gene="NNA1120" 69313.. .70134 /gene="NNA1120" /note="NNA1120, probable oxidoreductase, len: 273 aa; similar to many oxidoreductases e.g. TR:054197 (EMBL:AJ000671), car, Streptomyces clavuligerus clavulanate-9-aldehyde reductase (247 aa), fasta scores; E(): 2.1e-23, 36.6% identity in 238 aa overlap and SW:FABG.ECOLI (EMBL:M84991), fabg, Escherichia coli 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (244 aa), fasta scores; E(): 1.1e-16, 32.8% identity in 201 aa overlap. Contains Pfam match to entry PF00106 adh_short, short chain dehydrogenase and PS00061 Short-chain dehydrogenases/reductases family signature" /codon_start=1 /transl_table=11 /product="putative oxidoreductase" /protein_id="CAB84382.1" /db_xref="GI:7379814" /db_xref="SPTREMBL:Q9JUV3" /translation="MAVLITGASAGFGAMCRAFVIGAGYRVIGARRDLQALADEL GAFVPLEMDVSRRESVENALNGIDPEFSDIDCLLNAGLAGLDTADKADFEDNEM IQTNVLGLTFLTKILPQWVERGGYVMNLGSIAGNYATGNSVGTAKFVRQFSLN LRAELADNRVNTIEPGUCNTGTFNSVRFGKGDDEKAAGVYEGVFIREPDEIAETALW LYRFAHNNVNTIEPVAQTTFAGMKVIKAVPEVREDFEQSMLSLFSRSWFK" 69712.. .69798 /gene="NNA1120" /note="PS00061 Short-chain dehydrogenases/reductases family signature" 70138.. .70173 | RBS repeat_unit | complement(70638.. .70641) complement(70657.. .70811) /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 271.97" /label=Correia 70903.. .70906 70913.. .71845 /gene="pip" 70913.. .71845 /EC_number="3.4.11.5" /note="NNA1122, pip, probable proline iminopeptidase, len: 310 aa; similar to many e.g. SW:PIP_NEIGO (EMBL:Z25461), pip, Neisseria gonorrhoeae proline iminopeptidase (EC 3.4.11.5) (310 aa), fasta scores; E(): 0, 96.5% identity in 310 aa overlap. Contains Pfam match to entry PF00561 abhydrolase, alpha/beta hydrolase fold" /codon_start=1 /transl_table=11 /product="putative proline iminopeptidase" /protein_id="CAB84384.1" /db_xref="GI:7379816" /db_xref="SPTREMBL:Q9JUV1" /translation="MYETIKQPFHSGYQVSEIHQIYWEESGNPDGVPVIFLHGGPGAG ASPECGFENPDVFRIVIIDQRCGRSRPYACAEDNTWDLVADIEKVMELGIGKWL VFGSGWSTLSLAYQTHPERVKGLVLRGIFLCRPSETVNLNEAGSVRIYPEOMQKF VAPIAERNQLIEAYHGLLHQDEVCILSAKANADWESYLIRFEPEVEDADVASL ATARLENHYFYNGWLOGDRAILNINIGKIQHTPIIVQGRYDLCTPFMQSAWALSRAFP EAELRVQAGHRAFDPPLVDALVQAVEDILPLHL" 71084.. .71839 /gene="pip" /note="Pfam match to entry PF00561 abhydrolase, alpha/beta hydrolase fold, score 164.40, E-value 1.9e-45" complement(71902.. .73029) /gene="NNA1123" complement(71902.. .73029) /gene="NNA1123" /note="NNA1123, possible lipoprotein, len: 375 aa; unknown, contains an appropriately positioned PS00013 prokaryotic membrane lipoprotein lipid attachment site and a possible N-terminal signal sequence" /codon_start=1 /transl_table=11 /product="putative lipoprotein" /protein_id="CAB84385.1" /db_xref="GI:7379817" /db_xref="SPTREMBL:Q9JUV0" /translation="MTHIKPVIATAALIGLAACSGSKTEQPKLDYQSRSHRLIKLEVP PLNNPDGNYLRLPAGSGAVRASLDLEKRRTPAVOOPADAAYLVKSVKGYRLRDGSGR WLVPDGKSHAEITWPLLLKAFQENGPDIKSEPAIGOMETAEAKRAKIPQDSLRLRLFD TVGLGGYSTGERDKFIVRIEKGKNGVSDIFFAHRAMKEVYGGKDKDTTVWQPSPSDP NLEAFLTRFMQYLVGVGQQAENASAKPTLPAAENAMARIEGSLIVFDIGYRNNRPT ALALDRIGLTVVYQNTREHAFVLQAPNESNAVTEQKPLFKRLLLGKGAERKPAEQPE LVVYAEPAVANGSRIVLLNKGDSAYAGKDAALLGLHLSLR" complement(72973.. .73005) /gene="NNA1123" /note="PS00013 prokaryotic membrane lipoprotein lipid attachment site" complement(73037.. .73040) 73069.. .73101 /note="Stem loop containing DNA uptake sequences: gttat gcggtccgaa cgg ttcgacggc atggc" complement(73107.. .73982) /gene="dapA" complement(73107.. .73982) /gene="dapA" /EC_number="4.2.1.52" /note="NNA1124, dapA, probable dihydrodipicolinate synthase, len: 291 aa; similar to many e.g. SW:DAPA_ECOLI (EMBL:M12844), dapA, Escherichia coli dihydrodipicolinate synthase (EC 4.2.1.52) (292 aa), fasta scores; E(): 0, 49.3% identity in 292 aa overlap. Contains Pfam match to entry PF00701 DHDPS, Dihydrodipicolinate synthetase |
| misc_feature | 69313.. .69861 /gene="NNA1120" /note="Pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 191.70, E-value 1.2e-53" 69313.. .70134 /gene="NNA1120" 69313.. .70134 /gene="NNA1120" /note="NNA1120, probable oxidoreductase, len: 273 aa; similar to many oxidoreductases e.g. TR:054197 (EMBL:AJ000671), car, Streptomyces clavuligerus clavulanate-9-aldehyde reductase (247 aa), fasta scores; E(): 2.1e-23, 36.6% identity in 238 aa overlap and SW:FABG.ECOLI (EMBL:M84991), fabg, Escherichia coli 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (244 aa), fasta scores; E(): 1.1e-16, 32.8% identity in 201 aa overlap. Contains Pfam match to entry PF00106 adh_short, short chain dehydrogenase and PS00061 Short-chain dehydrogenases/reductases family signature" /codon_start=1 /transl_table=11 /product="putative oxidoreductase" /protein_id="CAB84382.1" /db_xref="GI:7379814" /db_xref="SPTREMBL:Q9JUV3" /translation="MAVLITGASAGFGAMCRAFVIGAGYRVIGARRDLQALADEL GAFVPLEMDVSRRESVENALNGIDPEFSDIDCLLNAGLAGLDTADKADFEDNEM IQTNVLGLTFLTKILPQWVERGGYVMNLGSIAGNYATGNSVGTAKFVRQFSLN LRAELADNRVNTIEPGUCNTGTFNSVRFGKGDDEKAAGVYEGVFIREPDEIAETALW LYRFAHNNVNTIEPVAQTTFAGMKVIKAVPEVREDFEQSMLSLFSRSWFK" 69712.. .69798 /gene="NNA1120" /note="PS00061 Short-chain dehydrogenases/reductases family signature" 70138.. .70173 | RBS stem_loop | complement(70159.. .70168) /note="Core DNA uptake sequence: ggcgtctgaa" /label=DUS complement(70185.. .70631) /gene="NNA1121" complement(70185.. .70631) /gene="NNA1121" /note="NNA1121, possible acyl-CoA hydrolase, len: 148 aa; shows weak similarity to part of SW:C7E2.RAT (EMBL:U49694) Rattus norvegicus cytosolic acyl coenzyme A thioester hydrolase (EC 3.1.2.2) (338 aa), fasta scores; E(): 4.2e-07, 31.0% identity in 113 aa overlap. Also similar to hypothetical proteins e.g. SW:YCIA_HAEIN (EMBL:U32765), HT0827, Haemophilus influenzae hypothetical protein (154 aa), fasta scores; E(): 0, 62.1% identity in 132 aa overlap. Similar to NMA1691, fasta scores; E(): 1.2e-08, 27.6% identity in 127 aa overlap. Contains Pfam match to entry PF01662 Acyl-CoA_hydro, Cytosolic long-chain acyl-CoA thioester hydrolase" /codon_start=1 /transl_table=11 /product="putative acyl-CoA hydrolase" /protein_id="CAB84383.1" /db_xref="GI:7379815" /db_xref="SPTREMBL:Q9JUV2" /translation="MQHEEGNRDPQGEILLRTVAMPRTNPNDIFGWMISOMDLG GGILAAETARGRIVTVAQEMNFIIRPVKGVNVCCYHCVRVGNLSLQKVEVWKTL MNDCTEDRHLYTEAVFTYVAIDACGNPRPIPKENPILAGLLPTP" complement(70212.. .70631) /gene="NNA1121" /note="Pfam match to entry PF01662 Acyl-CoA_hydro, Cytosolic long-chain acyl-CoA thioester hydrolase, score 160.80, E-value 2.3e-44" |
| misc_feature | 69313.. .69861 /gene="NNA1120" /note="Pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 191.70, E-value 1.2e-53" 69313.. .70134 /gene="NNA1120" 69313.. .70134 /gene="NNA1120" /note="NNA1120, probable oxidoreductase, len: 273 aa; similar to many oxidoreductases e.g. TR:054197 (EMBL:AJ000671), car, Streptomyces clavuligerus clavulanate-9-aldehyde reductase (247 aa), fasta scores; E(): 2.1e-23, 36.6% identity in 238 aa overlap and SW:FABG.ECOLI (EMBL:M84991), fabg, Escherichia coli 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (244 aa), fasta scores; E(): 1.1e-16, 32.8% identity in 201 aa overlap. Contains Pfam match to entry PF00106 adh_short, short chain dehydrogenase and PS00061 Short-chain dehydrogenases/reductases family signature" /codon_start=1 /transl_table=11 /product="putative oxidoreductase" /protein_id="CAB84382.1" /db_xref="GI:7379814" /db_xref="SPTREMBL:Q9JUV3" /translation="MAVLITGASAGFGAMCRAFVIGAGYRVIGARRDLQALADEL GAFVPLEMDVSRRESVENALNGIDPEFSDIDCLLNAGLAGLDTADKADFEDNEM IQTNVLGLTFLTKILPQWVERGGYVMNLGSIAGNYATGNSVGTAKFVRQFSLN LRAELADNRVNTIEPGUCNTGTFNSVRFGKGDDEKAAGVYEGVFIREPDEIAETALW LYRFAHNNVNTIEPVAQTTFAGMKVIKAVPEVREDFEQSMLSLFSRSWFK" 69712.. .69798 /gene="NNA1120" /note="PS00061 Short-chain dehydrogenases/reductases family signature" 70138.. .70173 | RBS gene | complement(70638.. .70641) complement(70657.. .70811) /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 271.97" /label=Correia 70903.. .70906 70913.. .71845 /gene="pip" 70913.. .71845 /EC_number="3.4.11.5" /note="NNA1122, pip, probable proline iminopeptidase, len: 310 aa; similar to many e.g. SW:PIP_NEIGO (EMBL:Z25461), pip, Neisseria gonorrhoeae proline iminopeptidase (EC 3.4.11.5) (310 aa), fasta scores; E(): 0, 96.5% identity in 310 aa overlap. Contains Pfam match to entry PF00561 abhydrolase, alpha/beta hydrolase fold" /codon_start=1 /transl_table=11 /product="putative proline iminopeptidase" /protein_id="CAB84384.1" /db_xref="GI:7379816" /db_xref="SPTREMBL:Q9JUV1" /translation="MYETIKQPFHSGYQVSEIHQIYWEESGNPDGVPVIFLHGGPGAG ASPECGFENPDVFRIVIIDQRCGRSRPYACAEDNTWDLVADIEKVMELGIGKWL VFGSGWSTLSLAYQTHPERVKGLVLRGIFLCRPSETVNLNEAGSVRIYPEOMQKF VAPIAERNQLIEAYHGLLHQDEVCILSAKANADWESYLIRFEPEVEDADVASL ATARLENHYFYNGWLOGDRAILNINIGKIQHTPIIVQGRYDLCTPFMQSAWALSRAFP EAELRVQAGHRAFDPPLVDALVQAVEDILPLHL" 71084.. .71839 /gene="pip" /note="Pfam match to entry PF00561 abhydrolase, alpha/beta hydrolase fold, score 164.40, E-value 1.9e-45" complement(71902.. .73029) /gene="NNA1123" complement(71902.. .73029) /gene="NNA1123" /note="NNA1123, possible lipoprotein, len: 375 aa; unknown, contains an appropriately positioned PS00013 prokaryotic membrane lipoprotein lipid attachment site and a possible N-terminal signal sequence" /codon_start=1 /transl_table=11 /product="putative lipoprotein" /protein_id="CAB84385.1" /db_xref="GI:7379817" /db_xref="SPTREMBL:Q9JUV0" /translation="MTHIKPVIATAALIGLAACSGSKTEQPKLDYQSRSHRLIKLEVP PLNNPDGNYLRLPAGSGAVRASLDLEKRRTPAVOOPADAAYLVKSVKGYRLRDGSGR WLVPDGKSHAEITWPLLLKAFQENGPDIKSEPAIGOMETAEAKRAKIPQDSLRLRLFD TVGLGGYSTGERDKFIVRIEKGKNGVSDIFFAHRAMKEVYGGKDKDTTVWQPSPSDP NLEAFLTRFMQYLVGVGQQAENASAKPTLPAAENAMARIEGSLIVFDIGYRNNRPT ALALDRIGLTVVYQNTREHAFVLQAPNESNAVTEQKPLFKRLLLGKGAERKPAEQPE LVVYAEPAVANGSRIVLLNKGDSAYAGKDAALLGLHLSLR" complement(72973.. .73005) /gene="NNA1123" /note="PS00013 prokaryotic membrane lipoprotein lipid attachment site" complement(73037.. .73040) 73069.. .73101 /note="Stem loop containing DNA uptake sequences: gttat gcggtccgaa cgg ttcgacggc atggc" complement(73107.. .73982) /gene="dapA" complement(73107.. .73982) /gene="dapA" /EC_number="4.2.1.52" /note="NNA1124, dapA, probable dihydrodipicolinate synthase, len: 291 aa; similar to many e.g. SW:DAPA_ECOLI (EMBL:M12844), dapA, Escherichia coli dihydrodipicolinate synthase (EC 4.2.1.52) (292 aa), fasta scores; E(): 0, 49.3% identity in 292 aa overlap. Contains Pfam match to entry PF00701 DHDPS, Dihydrodipicolinate synthetase |

```

/note="94% identity to consensus
CCAGTCGCTCAGTTTCAGTTCAGTTCGATAAATTCCTGTGCTTTCATTTCTAG"
/label=RS17
64513..64532
/note=">= 90% match to ATTCCNNNNNNNGGGAAT"
/label=DRS3
complement(64533..64558)
/note="98% identity to consensus
AAGACCGTCGGGCATCTCCAGCCGTC"
64559..64707
/note="Correia element; hmms hit to HMM Correia (1 -
151), score: 253.72"
/label=Correia
64749..64787
/note="Stem loop containing DNA uptake sequences: ctaat
gcgcgtctgaa cgacgaatg ttacagcgc atttt"
64754..64763
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(64773..64782)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(64799..67024)
/gene="icd"
complement(64799..67024)
/gene="icd"
/EC number="1.1.1.42"
/note="NMA1116, icd, probable isocitrate dehydrogenase,
len: 741 aa; similar to many e.g. SW:IDH_AZOVI
(EMBL:D73443), icd, Azotobacter vinelandii isocitrate
dehydrogenase (EC 1.1.1.42) (741 aa), fasta scores; E():
0, 65.6% identity in 735 aa overlap"
/codon_start=1
/transl_table=11
/product="putative isocitrate dehydrogenase"
/protein_id="CAB84378.1"
/db_xref="GI:7379810"
/db_xref="SPTREMBL:Q9JUV7"
/translation="MTQKSTIVVTHTEAPALATQSLPIVQAFARHADIDVKTADIS
LSGRILAAPPYEYTEAQRPDPAELGELVKQPDANVIKLPNISASVQPLTAIKELQ
SKGFAPVDXPADQTDDEKAVREYDRIGKSAVNPVLRGNSRRRAFKVAKFNKP
HSGATKDSKTHVATWQSGDFHNEQSVTPDTSVIVFTDKQGNKKELRAPVAMK
AGBIDATVMSKALLAFIAEQVDAAKAGVFLSLHMKATMMKVSDPIIFGHAVKAF
APVFERGKLAAGVNVNNGFNLLNDKLDADTRAAVEAETAAVYANPDLAMVD
SDKITNLHVPSDVIDASMPAMIRNSGRWDKGAQDTKAVIPDSSVAGVYQATID
FCREHGFDPDTMTGTPNVGLMAQAAEYGSNKTFEIEADGOVQVIDAAGKVLMOHD
VEAGDIWRMCTKDPVKDMVOLAVNRTRLSNTPAVFWLDENRPHKSLKAYLA
ELDTNGLDIRVLAPESAAKPSGLRNGEDTISVTGNVLRDYLDFPPIELGTSAM
LSIVPLMNGGMPETGAGGSAKPVQOOFLEENHLRWDLSLGEFLAIVSEFHLAQKTN
AKQVLADTLDAATEKLLNDKSPKRKAGELDNRGSHFYLTLYWAOELAAQDKDAELK
AAFTPLAALTADEAKIVAELSAVQGAADIGGYAANPEKAAQVMRPSVTENQALNA
L"
67057..67254
/gene="NMA1117"
67057..67254
/note="NMA1117"
/note="NMA1117, len: 65 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1117"
/protein_id="CAB84379.1"
/db_xref="GI:7379811"
/db_xref="SPTREMBL:Q9JUV6"
/translation="MEHIARGMCAIMAYFGGFVRFALFDLGVFERGVKGRKWFSSRV
WRRSEVLNMGCGGIFSVKLIIL"
complement(67295..67356)
/note="Correia element; hmms hit to HMM Correia (1 - 61),
score: 62.95"
/label=Correia
67443..67446
67457..68572
/gene="lst"
67457..68572

```

```

/gene="lst"
/EC number="2.4.99.-"
/note="NMA1118, lst,
CMP-N-acetylneuraminate-beta-galactosamide-
alpha-2,3-sialyltransferase, len: 371 aa; highly similar
to sialyltransferases from other strains e.g. TR:P72097
(EMBL:U60661), lst, Neisseria meningitidis strains 406Y,
NRCC and 4030 CMP-N-acetylneuraminate-beta-galactosamide-
alpha-2,3-sialyltransferase (EC 2.4.99.-) (371 aa), fasta
scores; E(): 0, 98.9% identity in 371 aa overlap"
/codon_start=1
/transl_table=11
/product="CMP-N-acetylneuraminate-beta-galactosamide-
alpha-2,3-sialyltransferase"
/protein_id="CAB84380.1"
/db_xref="GI:7379812"
/db_xref="SPTREMBL:Q9JUV5"
/translation="MGLKACLTVLCILVFCFGIFYTDFRVNOCERNASVLLKDKLFN
EEGEPVNLIFCYTILQMKVAERIMAOHPGRFVYVLMSENREKDYVFQIKDKAER
AYFPHLPYGLNKSNFPIPTMAELKVKSMLLPKYKRIYLALEKVSIAAFLSTYPDARI
KTDDGNGNLIQSSSYLCDEFVNGTIRKRFARMIGDWSIAKTRNASDEHYTIFKGL
KNIMDGRKRMVYLPFLFDASLKGADGTGTVRILLGSPDKMEKEISEKAAKNFNIQY
VAPHPRTYGLSGVLTNTPYVIEDYILREIKNPHTRYEIYTFPSGAALTMKDFPNV
HVYALKPASLPEDYWLKXPVVALFTQSGIPILTFDDKD"
68690..68722
/note="Stem loop containing DNA uptake sequences: aaaat
gcgcgtctgaa ggc ttacagcgc attgt"
68695..68704
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(68708..68717)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(68736..69194)
/gene="NMA1119"
complement(68736..69194)
/gene="NMA1119"
/note="NMA1119, probable C-type cytochrome, len: 152 aa;
similar to many e.g. SW:C554_RHOSH (EMBL:L36880), cyfC,
Rhodobacter sphaeroides cytochrome C-554 precursor (153
aa), fasta scores; E(): 9.9e-06, 28.4% identity in 148 aa
overlap. Contains Pfam match to entry PF01322
Cytochrome_C_2, Cytochrome C', PS00190 Cytochrome c family
heme-binding site signature and an appropriately
positioned PS00013 Prokaryotic membrane lipoprotein lipid
attachment site. Contains a probable N-terminal signal
sequence"
/codon_start=1
/transl_table=11
/product="putative C-type cytochrome"
/protein_id="CAB84381.1"
/db_xref="GI:7379813"
/db_xref="SPTREMBL:Q9JUV4"
/translation="MRTOISLAAATLTLISACGSGIPSQPKGEISENRTAAFKSNM
PDFSRMGKMWKGEEDYDFEKFQAAAAFAESSKKPTLFESDPQGNRALPAVWSDGA
KFEAEKTKFAAAVEKLNAAQTGKLDKRAAYGETCASCSDSFRAPF"
complement(68742..69107)
/gene="NMA1119"
/note="Pfam match to entry PF01322 Cytochrome_C_2,
Cytochrome C', score 113.40, E-value 4.2e-30"
complement(68757..68774)
/gene="NMA1119"
/note="PS00190 Cytochrome c family heme-binding site
signature"
complement(69138..69170)
/gene="NMA1119"
/note="PS00013 prokaryotic membrane lipoprotein lipid
attachment site"
complement(69205..69208)
69250..69259
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
69303..69306
RBS

```

identity in 399 aa overlap. Shows weak similarity to
NMA1684, fasta scores: E(): 3.4e-11, 23.8% identity in 349
aa overlap. Contains Pfam match to entry PF00155
aminotran_1, Amino transferases class-1
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="CAB84375.1"
/db_xref="GI:7379807"
/db_xref="SPTREMBL:Q9JUW0"
/translation="MNTLLKQLKPYFFARLHEAMQGISAPEGMEAVPLHIGEPKHPTP
KYITDALTSLEHEKYPLTAGLPQLACANMLRRYDGLTVADADNEILPVLGREA
LFSFQTVLNPVSDGKPAIVSPNPFYQIYEAGTLIGGEEIHFAISDCPAPSPNPDWRSI
SEEVKRRKLVEVCPNPNPSGVLDDGWKEVFDLQDKYGFIIASDECYSEIYFDGNK
PLGCLQAQLGRSRQKLLMF"SLSKRSNVPLGRSGFVAGDAELLKNELLYRTYHGS
MSIPQVRSIAAWDEQHVINDNRRLYQEKFERVIPILQVDFVKLPDASFYIWLKVPD
GDGLAFARNLWOKAAIQVLPGRFLARDTEQGNPGEYRIALVADVATCVKAAEDIVS
LYR"
59755..60915
/gene="NMA1113"
/note="Pfam match to entry PF00155 aminotran_1,
Amino transferases class-1, score 8.50, E-value 2.5e-10"
misc_feature
complement(60072..60081)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
60926..60961
/note="Stem loop containing DNA uptake sequences: gaaat
gccgtctgaa ctttgg ttacagacggc atttt"
60931..60940
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(60947..60956)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
60998..61059
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 95.11"
/label=Correia
61061..61103
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 50.38"
/label=Correia
61180..61959
/gene="NMA1114"
61180..61959
/gene="NMA1114"
/note="NMA1114", len: 259 aa; similar to bacterial
hypothetical proteins e.g. SW:YAAA_ECOLI (EMBL:D10483),
yaaA, Escherichia coli hypothetical protein (258 aa),
fasta scores: E(): 0, 54.9% identity in 257 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1114"
/protein_id="CAB84376.1"
/db_xref="GI:7379808"
/db_xref="SPTREMBL:Q9JUW9"
/translation="MEFVLSPAKLNKEDPAPVSEFTOPDLLAESDILLMOQLRELAPQ
QIAELMHVSDKJALLNAORNAEWNPFTPENAKQAVFMFGDVEGMNANTLNTQIQ
YIQGRVLLSGLYGLLRPLDLIOPVRLMGTFANLRGNLYEPFGDIIITNLNDTLA
QAGSTNLVASQELFYKSVNTKKRLRLITPIFKDEKNGKYKIIISFYAKRARGLMVRY
AAEHNITDPEMLKFNFEYAFNDAASNESEWFMRSQIK"
62042..62132
/note="tRNA Ser anticodon TGA, Cove score 69.59"
repeat_unit
complement(62201..62338)
/note="REP 2; hmms hit to HMM REP 2 (1 - 128), score:
150.65"
tRNA
62345..62434
/note="tRNA Ser anticodon CGA, Cove score 64.96"
62483..63698
/note="IS1106A3"
/label=IS1106A3
62483..62500
/note="inverted repeat at end of IS1106A3"

/label=IR
complement(62630..63637)
/gene="NMA1115"
complement(62630..63637)
/gene="NMA1115"
/note="NMA1115, probable transposase for IS1106A3, len:
335 aa; similar to many e.g. TR:CAB44967 (EMBL:AJ242841),
Neisseria meningitidis transposase for insertion element
IS1106A3 (335 aa), fasta scores: E(): 0, 97.6% identity in
335 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transposase for IS1106A3"
/protein_id="CAB84377.1"
/db_xref="GI:7379809"
/db_xref="SPTREMBL:Q9JUW8"
/translation="MSTFRFQAQAMIAKHIDRFPLLKLDQVIDMPOIEQYLNLRQTR
YLRDRHRGPAYPLLSMFRVAVLGGHMSDPELHSLITRIDFNFCRDELSIPDYS
TLCRYRNWLAQDDTLSELKLNLCOLTEKGLIKERASAAVVDATIIQTAGSKQROAIE
VDEEGYISQOTTPSKSDARWIKKGLYKGYKQHTDAGYIEKLYITPTNAHECT
HLSPLEGLPKGTTYADKGVDSAEHQHLEHLODGMKACNRPLSEVQTKRNR
YLSKTRYVVEQSFGLHRRFRYARAYFGLIKVSAQSHLKACMLNLLKANRLSAPAA
A"
complement(63681..63698)
/note="inverted repeat at end of IS1106A3"
/label=IR
63819..63973
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 258.02"
/label=Correia
complement(63974..63993)
/note="100% identity to consensus GACGCGACAGGGTTGCTGT"
63994..64013
/note=">= 90% match to ATTCCNNNNNNNNGGAAT"
/label=DRS3
64014..64069
/note="99% identity to consensus
CCAGTCGCTCAGTTTCAGTCATTCGATAAATCCTTGTCTTTTCATTTCTAG"
/label=RS17
64070..64089
/note=">= 90% match to ATTCCNNNNNNNNGGAAT"
/label=DRS3
64090..64145
/note="97% identity to consensus
GACGCGAAGCGTTTGGTTTTCCGATAAATCTTCAGGCATTGAATTCAG"
/label=RS18
64146..64165
/note=">= 90% match to ATTCCNNNNNNNNGGAAT"
/label=DRS3
64166..64235
/note="100% identity to consensus
GACGATTCATAGCTTCCGAAATTCACATAACCGAAACCTGACAGTAACCGTAG
CAACTGAACCGTC"
/label=RS16
64236..64255
/note=">= 90% match to ATTCCNNNNNNNNGGAAT"
/label=DRS3
64256..64311
/note="88% identity to consensus
CTAGATCTCGGACTTCAGATAATCTTTGAATATTTGTTTCTTAAGGCTAG"
/label=RS13
64312..64331
/note=">= 90% match to ATTCCNNNNNNNNGGAAT"
/label=DRS3
64332..64437
/note="94% identity to consensus
GACGATTCATCCGACGAAACCTGCACACGTCATTCTACGAACCTACATCCCG
TCATTCGCCAAGGACAGAAACCAATCAGAAACCTTAAATCCCGTC"
/label=RS14
64437..64456
/note=">= 90% match to ATTCCNNNNNNNNGGAAT"
/label=DRS3
complement(64457..64512)

| | | | |
|-------------|---|--------------|---|
| repeat_unit | 57729..57787 /note="97% identity to consensus GACGAATCCATCCGTACCGAAACCTGCACACGTCATTCCACGAACCTACATCCCG TC" | repeat_unit | /label=QRS3 58489..58643 /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 274.69" |
| repeat_unit | /label=RS14a 57788..57807 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | gene | /label=Correia complement(58578..59060) |
| repeat_unit | /label=QRS3 57808..57863 /note="98% identity to consensus CCAGTTTTCAGTTTCAGTCATTCCGATAAATTCCTTAGCATTTGAATGCTAG" | CDS | /gene="NMA1111" complement(58578..59060) |
| repeat_unit | /label=RS26 57864..57883 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | /note="NMA1111, len: 160 aa; shows weak similarity to N-terminal regions of bacterial hypothetical proteins e.g. SW:GAZ_ECOLI (EMBL:AE000353), ygaZ, Escherichia coli hypothetical protein (245 aa), fasta scores; E(): 1.7e-05, 28.4% identity in 116 aa overlap" |
| repeat_unit | /label=QRS3 57884..57942 /note="99% identity to consensus GACGAATCCATCCGTACCGAAACCTGCACACGTCATTCCACGAACCTACATCCCG TC" | | /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1111" |
| repeat_unit | /label=RS14a 57943..57962 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | /protein_id="CAB84373.1" |
| repeat_unit | /label=QRS3 57963..58018 /note="98% identity to consensus CCAGTTTTCAGTTTCAGTCATTCCGATAAATTCCTTAGCATTTGAATGCTAG" | | /db_xref="GI:7379805" |
| repeat_unit | /label=RS26 58019..58038 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | RBS | /db_xref="SPTREMBL:Q9JUW2" |
| repeat_unit | /label=QRS3 58038..58093 /note="99% identity to consensus TGACGGATTTAGGTGGGGCATTATTGGAAAAAGCAGAAACCGCTCCGCCGTC" | gene | /translation="MTHISSPRNEFIRGKISSPMLIGLLPWALILGMOGGKGMSP EMLLMTGMNFAGSEFATVNLWAEPLILLIATVTFMINSRHILMGAALAPHLKEIPL KKAVPALFFMCDSEWAMAFSEIQKKAAGLPAFNPFPYSGLTSTALPRLSSKRTIL " |
| repeat_unit | /label=RS30 58094..58113 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | 59191..59195 59203..59655 /gene="dut" |
| repeat_unit | complement(58114..58173) /note="91% identity to consensus CCAGACCTGTCGCGACGGAACCTTATCGGATAAAACGGTTTCTCCGATCCTACGTT TAG" | CDS | 59203..59655 /gene="dut" |
| repeat_unit | /label=RS35 58174..58193 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | /EC_number="3.6.1.23" |
| repeat_unit | /label=QRS3 58194..58252 /note="97% identity to consensus GACGAATCCATCCGTACCGAAACCTGCACACGTCATTCCACGAACCTACATCCCG TC" | | /note="NMA112, dut, probable deoxyuridine 5'-triphosphate nucleotidohydrolase, len: 150 aa; similar to many e.g. SW:DUT_ECOLI (EMBL:X01714), dut, Escherichia coli deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23) (151 aa), fasta scores; E(): 0, 65.3% identity in 147 aa overlap. Contains Pfam match to entry PF00692 dUTPase, dUTPase" |
| repeat_unit | /label=RS14a 58253..58272 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | /codon_start=1 /transl_table=11 /product="putative deoxyuridine 5'-triphosphate nucleotidohydrolase" |
| repeat_unit | /label=QRS3 58273..58328 /note="98% identity to consensus CCAGTTTTCAGTTTCAGTCATTCCGATAAATTCCTTAGCATTTGAATGCTAG" | misc_feature | /protein_id="CAB84374.1" |
| repeat_unit | /label=RS26 58329..58348 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | /db_xref="GI:7379806" |
| repeat_unit | /label=QRS3 58349..58383 /note="94% identity to consensus GACGAATCCATCCGTACCGAAACCTGCACACGTC" | stem_loop | /translation="MTEVEKVLDERMADVVPVYATKGSAGLDLRACLDDEVVLOPG ETELVPTGLATYLDAPSYAAVLLPRSGLGHKHGIVLGLIDSDYOGELKVSILNWR SSEPTVKKPERIAQMWIVPIVQARFRKRVVEFVSSRGEGGFGSTGLH" |
| repeat_unit | /label=RS14b 58384..58403 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | misc_feature | 59242..59646 /gene="dut" |
| repeat_unit | /label=QRS3 58404..58459 /note="98% identity to consensus CCAGTTTTCAGTTTCAGTCATTCCGATAAATTCCTTAGCATTTGAATGCTAG" | | /note="Pfam match to entry PF00692 dUTPase, dUTPase, score 158.80, E-value 9.6e-44" |
| repeat_unit | /label=RS26 58460..58479 /note=">= 90% match to ATTCCNNNNNNNGGGAAT" | | 59660..59701 /note="Stem loop containing DNA uptake sequences: aaaat ggcgctctgaa gggcagctcagg ttcagacgc atata" |
| | | misc_feature | /note="Core DNA uptake sequence: gccgctctgaa" |
| | | | /label=DUS complement(59687..59696) |
| | | | /note="Core DNA uptake sequence: gccgctctgaa" |
| | | RBS | /label=DUS 59721..59724 |
| | | gene | 59731..60918 /gene="NMA1113" |
| | | CDS | 59731..60918 /gene="NMA1113" |
| | | | /note="NMA113, probable aminotransferase, len: 395 aa; shows weak similarity to SW:AAT_BACSP (EMBL:M59430) Bacillus sp. (strain YM-2) aspartate aminotransferase (EC 2.6.1.1) (392 aa), fasta scores; E(): 3.8e-21, 24.9% identity in 366 aa overlap. Similar to many putative aminotransferases e.g. TR:Q9ZEX3 (EMBL:A009834), dapC, Bordetella pertussis putative succinylidaminopimelate transaminase (397 aa), fasta scores; E(): 0, 52.6% |

repeat_unit

```
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
53732. .53887
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 264.87"
/label=Correia
53914. .54579
/gene="NMA1106"
53914. .54579
/gene="NMA1106"
/note="NMA1106, len: 221 aa; unknown, N-terminus is
similar to the N-termini of fibrial proteins e.g.
TR:051528 (EMBL:L48934), fimt, pseudomonas aeruginosa
fibrial protein (169 aa), fasta scores; E(): 0.00046,
31.5% identity in 108 aa overlap. Contains possible
N-terminal signal sequence, similar to that present in
others from Neisseria meningitidis: NMA0264, NMA110,
NMA0726, NMA0424, which are all pilin-related"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1106"
/protein_id="CAB84368.1"
/db_xref="GI:7379800"
/db_xref="SPTREMBL:Q9JUW7"
/translacion="MCTRKQGGFTLTLLVIMVIAAYMAMIALPNMSQWIASRRRIASH
AERIANLRRSGEAVRLNLPVYICPVQVKKDGTVNNKCDGSKGQGLAFAGDKNGNK
TYDGAADVFLRSVVLNDDSDSRINAFNHFAGSSOPTADRVVVTENQGTGTYTG
QHLTKSSFPYSDCYIQIIVLTDARAVSDADKKFSAVVLNSSGRVEVCPNDRRTVC
QYK"
```

gene

54609. .55223

CDS

```
/gene="NMA1107"
54609. .55223
/gene="NMA1107"
/note="NMA1107, possible membrane protein, len: 204 aa;
unknown, N-terminus is similar to the N-terminus of
TR:057003 (EMBL:L36117), pilv, Pseudomonas aeruginosa gene
required for type 4 fibrial biogenesis (185 aa), fasta
scores; E(): 0.00046, 25.1% identity in 183 aa overlap.
Contains hydrophobic, possible membrane-spanning region
near the N-terminus"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84369.1"
/db_xref="GI:7379801"
/db_xref="SPTREMBL:Q9JUW6"
/translacion="MKNNDCFRLNKPSQGMALIEVLVAMLVLTIGILALLSVQLRTVA
SVRAETQIVSQITQNLMEGLMNPITSDSNKKNLYNMGHHLASVVDGDFQVDA
IKTKQLAEQLKRFSEYELKNALPDAAAIHYAVCKDSSGVAPTLISAGSTFSSNCDGSA
NGDTLIRLVNPNDSAGSDIARTNLETNGNNIVTYQARVGGRE"
```

RBS

55209. .55213

gene

55220. .56218

CDS

55220. .56218

```
/gene="NMA1108"
55220. .56218
/gene="NMA1108"
/note="NMA1108, possible membrane protein, len: 332 aa;
unknown, contains hydrophobic, possible membrane-spanning
region near the N-terminus"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84370.1"
/db_xref="GI:7379802"
/db_xref="SPTREMBL:Q9JUW5"
/translacion="MRRKMLNVPKNGYDGMKGFTIIEFLVAGMLSMITVLMVAGSSYPT
SRKLNDAANRLSAQDLRNAATLIVDRMAGGFGCFNMSSEHTKNDIIVDPKQTOH
VPVPGAKQENPLFSEKWIANTNNNTAKLIPIAESTDIKYPGFAOARPALIFOYGI
DLDASAEIVVSSCSKIAPGKKISTLOPAKSALQITNDKONGNTTORHVVNAYA
VGRTAGEGFLRFOLDKKGWNPOLLVKKIRHMKVRYIYSDCPEDDDAGKEKPKY
TGTFDSSTNAVTPAGVEVLSSGTDTKIAASDNHHIYRIDATIRGNVNCANRTL"
```

gene

56197. .56790

CDS

56197. .56790

```
/gene="NMA1109"
/note="NMA1109, possible membrane protein, len: 197 aa;
unknown, contains hydrophobic, possible membrane-spanning
region near the N-terminus"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84371.1"
/db_xref="GI:7379803"
/db_xref="SPTREMBL:Q9JUW4"
/translacion="MRKONTLTGPTSDGQGFALFIVLMVIMVIVFVLVTAASQSYNT
EQRISANESDRKLASLAAREGLAQVLDLEYDTSKVTFSENCQGLCTAVNVRT
NNDNEAFDNIVQKPTVEAVKRCSKTAKSTGLCIDNKGMEYKKGKTSQSVKMPRYIE
YLVGKGENYVRTAKAWGNANTVWVVLQSVSNDE"
56780. .57268
/gene="NMA1110"
56780. .57268
/gene="NMA1110"
/note="NMA1110, probable pilin, len: 162 aa; similar to
many e.g. SW:EMI_MORBO (EMBL:M32345) Moraxella bovis
fimbrial protein I precursor (159 aa), fasta scores; E():
6.4e-06, 28.1% identity in 153 aa overlap. Contains Pfam
match to entry PF00114 pilin, Pilins (bacterial filaments)
and PS00409 Prokaryotic N-terminal methylation site.
Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative pilin"
/protein_id="CAB84372.1"
/db_xref="GI:7379804"
/db_xref="SPTREMBL:Q9JUW3"
/translacion="MMSNMKEQKGTLLIEMMVVAILGIISVIAIPSVQSVIEKGYOS
OLYTEWGINNISQFILKNPLDNDQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNE
EKPRAYSLVGVPKFTGTITLSVMNNSVGDGFKCDRAASRAHLETLSDVCEAFSNN
KK"
56804. .56866
/gene="NMA1110"
/note="PS00409 Prokaryotic N-terminal methylation site"
56810. .57247
/gene="NMA1110"
/note="Pfam match to entry PF00114 pilin, Pilins
(bacterial filaments), score -15.90, E-value 0.001"
57284. .57293
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
57323. .57477
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 278.82"
/label=Correia
57487. .57506
/note=">= 90% match to ATTCNNNNNNNNNGGAAAT"
/label=dRS3
57507. .57560
/note="96% identity to consensus
CTAGAAATTTAATGTTGCGGCACCTAGCCAAAAAACCGAAACCGACGGACTAG"
/label=RS29
57561. .57580
/note=">= 90% match to ATTCNNNNNNNNNGGAAAT"
/label=dRS3
57580. .57635
/note="99% identity to consensus
TGACGGATTTTAGTTGGGGCATTTATTGAAAAACAGAAACCGCTCCGCCGTC"
57636. .57654
/label=RS30
57655. .57708
/note="100% identity to consensus
CTAGAAATTTAATGTTGCGGCACCTAGCCAAAAAACCGAAACCGACGGACTAG"
/label=RS29
57709. .57728
/note=">= 90% match to ATTCNNNNNNNNNGGAAAT"
/label=dRS3
```


/protein_id="CAB84361.1"
/db_xref="GI:7379793"
/db_xref="SPTREMBL:O9JUX3"
/translation="MKPYSANPNLTERRRLRVLLIAALGELLMLVPLVAVPYEAL
KGMDDLKSLNDEPWSAKLITLITLIVPVNAVIGVAMWLLTRDFRQKLLLT
LLDLPFSVPVAGMLVLLFGAHTALGWLEAGQIQIIFAIPIGIIILATLFTVFPFYA
REIPLMAQAGDEQAALILGASGQMFWRTLPNIKWALLYGIIILTNARAMEFGA
VSVVSHIRGETNTVPLLVEIFEXNEYNFTGAFALSGVLLALLATLAVQNIIKLQDK
KLAABERNV"
complement(46672..46675)
/gene="cysw"
/note="cysw"
complement(46787..47026)
/gene="cysw"
/note="Pfam match to entry PF00528 BPD_transp,
Binding-protein-dependent transport systems inner membrane
component, score 39.50, E-value 7.8e-08"
complement(46937..47023)
/gene="cysw"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp sign."
47556..47598
/note="Stem loop containing DNA uptake sequences: ggag
gtcgtctgaa aaggattgagggt ttcagacgac cttt"
complement(47586..47774)
/gene="NMA1099"
complement(47586..47774)
/gene="NMA1099"
/note="NMA1099, len: 62 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1099"
/protein_id="CAB84362.1"
/db_xref="GI:7379794"
/db_xref="SPTREMBL:O9JUX2"
/translation="MTAARLFYFSFKDKRSSENASAPQTEFVGRIYSDKHLRFS
NLXGKYPTYLMNKKVY"
47773..47792
/note=">= 90% match to ATTCNNNNNNNGGGAAT"
/label-drs3
complement(47793..47850)
/note="RS99"
/label="RS99"
47851..47870
/note=">= 90% match to ATTCNNNNNNNGGGAAT"
/label-drs3
complement(47899..48735)
/gene="cysu"
complement(47899..48735)
/gene="cysu"
/note="NMA1100, cysu, probable sulphate permease inner
membrane protein, len: 278 aa; similar to many e.g.
SW:CYST_ECOLI (EMBL:M32101), cysu, Escherichia coli
sulfate transport system permease protein (277 aa), fasta
scores: E(): 0, 48.4% identity in 273 aa overlap. Contains
pfam match to entry PF00528 BPD_transp.
Binding-protein-dependent transport systems inner membrane
component and PS00402 Binding-protein-dependent transport
systems inner membrane comp sign. Contains hydrophobic,
probable membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative sulphate permease inner membrane
protein"
/protein_id="CAB84363.1"
/db_xref="GI:7379795"
/db_xref="SPTREMBL:O9JUX1"
/translation="MLAKTPGVLPFGKSLGLTGLVLSLVLPFAMMAAKAEIGW
SGFWNTITENVLAAVWLSLRMSFYAMLTVNVFGTLVAVWLVRYEPGKLANALVDL
PFALPFAVTGIALATLYAPNGWIGRFEPLGKIAETPVGINIALVVSVPFIVRAVQ
PVLELSGSETEAAALGSRWTFTRRVLLPEITPALLTGAGMWFARATGEYGSVIFI
AGNIPMVSELPILITCKLEQFDVQASAVLPMLLVSFVILPALINMQWALRRSGA
KG"
complement(48010..48252)

/gene="cysU"
/note="Pfam match to entry PF00528 BPD_transp,
Binding-protein-dependent transport systems inner membrane
component, score 60.40, E-value 3.9e-14"
complement(48163..48249)
/gene="cysU"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp sign."
48877..48939
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 74.65"
/label="Correia"
48941..48983
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 51.92"
/label="Correia"
49016..49019
49026..49358
/gene="NMA1101"
49026..49358
/gene="NMA1101"
/note="NMA1101, len: 110 aa; unknown, Glu-rich protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1101"
/protein_id="CAB84364.1"
/db_xref="GI:7379796"
/db_xref="SPTREMBL:O9JUX0"
/translation="MQILSFQPDIAERMLTEGESVNEAQFVRTONGYVIAHHEGV
AALLAPDTPPGIPCFWVEGAESLEELCVMERGEFDEVEFDGDDLEWLETAQCGHH
GDACACGH"
49690..50199
/gene="NMA1102"
49690..50199
/gene="NMA1102"
/note="NMA1102, possible integral membrane protein, len:
169 aa; similar to bacterial hypothetical proteins e.g.
SW:Y318_HAEIN (EMBL:U32717), HI0318, Haemophilus
influenzae (172 aa), fasta scores: E(): 0, 56.2% identity
in 169 aa overlap. Contains PS00017 ATP/GTP-binding site
motif A (P-loop). Contains hydrophobic, possible
membrane-spanning regions. Lies within a region of
unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84365.1"
/db_xref="GI:7379797"
/db_xref="SPTREMBL:O9JRI2"
/translation="MTMILSILSFFIIRLLFLAVSTIKHEKALIAKAKOYKGTNSTL
LAHVHLYLACFWVWLSDTAFNGISLIGTLFVMAISFVILSLIKOLGEIWTVKIIL
LPHQINRSWLFFKTRHPNFIILPELIGALLCQAWYLLIGLPIYLLVLFKRRIQ
EEQAMATLF"
49786..49809
/gene="NMA1102"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(50214..50255)
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 54.94"
/label="Correia"
50285..50344
/note="98% identity to consensus
TTCCGCCATCTCAAGATTTACAGCGATACACGGTGATTAAAGGAATGCCGAACC
GTC"
50345..50363
/note=">= 85% match to ATTCNNACNACNTTTCGTC"
/label="RS3x"
50364..50383
/note=">= 90% match to ATTCNNNNNNNGGGAAT"
/label-drs3
50384..50418
/note="100% identity to consensus
CTAGGACGCGAGGTTAAGAAACCTACATCCCGTC"


```
misc_feature complement(38642..38651)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
misc_feature complement(38961..38970)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
RBS 39075..39078
gene 39086..39667
/genes="NMA1091"
CDS 39086..39667
/genes="NMA1091"
/notes="NMA1091, possible outer membrane lipoprotein, len:
193 aa; shows weak similarity to bacterial outer membrane
lipoproteins e.g. SW:LOLB.PSEAE (EMBL:X82071), lolB,
Pseudomonas aeruginosa outer membrane lipoprotein, fasta
scores: E(): 3.5e-05, 26.9% identity in 197 aa overlap.
contains probable N-terminal signal sequence and
appropriately positioned PS00013 prokaryotic membrane
lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative outer membrane lipoprotein"
/protein_id="CAB84354.1"
/db_xref="GI:7379786"
/db_xref="SWISS-PROT:P57023"
/translation="MKHTVSASVILLITACAQLPQNNENLWOPSEHRSFTABGR LAV
KAEKGSYANFDWTYOPVETININTPLGSTGLQCQDRDGLAVDGKGNVYQAESAE
ELSRQLVGFKLPIQLHINWADGRVAGAPYRIIPDGLILEQYGVTVGRTADSGQVRTL
QLNNGNLINILVTEIGMPSETETQEQCAARIQ"
/label=DUS
misc_feature 39101..39133
/genes="NMA1091"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
39619..39628
misc_feature 39619..39628
/genes="NMA1091"
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
gene 39677..40522
/genes="NMA1092"
CDS 39677..40522
/genes="NMA1092"
/notes="NMA1092, len: 281 aa; similar to many hypothetical
proteins e.g. SW:YCHB.ECOLI (EMBL:M77237), ychB,
Escherichia coli hypothetical protein (283 aa), fasta
scores: E(): 0, 42.4% identity in 278 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1092"
/protein_id="CAB84355.1"
/db_xref="GI:7379787"
/db_xref="SPTREMBL:Q9JUX8"
/translation="MSVTGRQAFPAKLNLDLRITGRREDGYHHNIESIFCLIDLQD
TVYKPRDDGKILIHNPVDGMPQEAADLSYRAASLLQKYARTPTGVFIWLDKRIPTGAG
LGGSDDATVLLVLRNWMQCGILQRLIDSGAALGADVPFFTFGNKAFARGICDRLD
EMDIPKOWYVIVKPPVHVSTAKIFTHEGLTRNSASSIMPTFQNLQPRNDMOAVFKE
YPEWVKAYSLSRYGALMTGSGACVFTACQDRNSAYNIYRQVSDLYEAYLAESLSKH
PLLSV"
40530..40605
tRNA /notes="tRNA Gln anticodon TTG, Cove score 71.63"
40618..40693
tRNA /notes="tRNA Gln anticodon TTG, Cove score 71.63"
40702..40777
tRNA /notes="tRNA Gln anticodon TTG, Cove score 71.63"
40777..40814
stem_loop 40777..40814
gene 40944..41927
/genes="prsa"
CDS 40944..41927
/genes="prsa"
/EC_number="2.7.6.1"
/notes="NMA1093, prsA, probable ribose-phosphate
pyrophosphokinase, len: 327 aa; similar to many e.g.
SW:KPRS.ECOLI (EMBL:M13174), prsA, Escherichia coli
ribose-phosphate pyrophosphokinase (EC 2.7.6.1) (314 aa),
```

```
fasta scores; E(): 0, 66.5% identity in 310 aa overlap.
Contains pfam match to entry PF00492 PRPP_synth,
Phosphoribosyl pyrophosphate synthetase, PS00103
purine/pyrimidine phosphoribosyl transferases signature
and PS00114 Phosphoribosyl pyrophosphate synthetase
signature"
/codon_start=1
/transl_table=11
/product="putative ribose-phosphate pyrophosphokinase"
/protein_id="CAB84356.1"
/db_xref="GI:7379788"
/db_xref="SPTREMBL:Q9JQV4"
/translation="MAAYDSLWVLTGTNANPELAQRVVVRHLDISLGNASVKFSDEGEVA
VELLENVRGRDVFILQPTCAPTDNLMEILTMADALKRASAGRTTATPVGVAROOR
RPSVRVPISAKLVANMLYSAGIDRVLTVDLHADQIQGFEDIPVDNIYATPILINDIK
QQRLENLTVVSPIGIGVVRARAVAKSLNADLAIIDRRPKANVAEWNIIIGDQGRFC
LIVDDMIDTATLCKAAVKERVAERVLAYASHVFSGEAVSRKASSEIDQVVVTDT
IPLESAKNCDRIRQVITIAGLLAETVRRISNEESVSYLFNEEYVMTGSMLLP"
40962..41888
misc_feature 40962..41888
/genes="prsa"
/notes="Pfam match to entry PF00492 PRPP_synth,
Phosphoribosyl pyrophosphate synthetase, score 621.00,
E-value 6.8e-183"
complement(41057..41066)
misc_feature complement(41057..41066)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
41337..41384
misc_feature 41337..41384
/genes="prsa"
/notes="PS00114 Phosphoribosyl pyrophosphate synthetase
signature"
41595..41633
misc_feature 41595..41633
/genes="prsa"
/notes="PS00103 Purine/pyrimidine phosphoribosyl
transferases signature"
41982..41985
gene 41982..41985
/genes="NMA1094"
CDS 41994..42566
/genes="NMA1094"
41994..42566
/genes="NMA1094"
/notes="NMA1094, len: 190 aa; shows weak similarity to
SW:CTC_BACSU (EMBL:D26185), etc, Bacillus subtilis general
stress protein (203 aa), fasta scores: E(): 6.7e-11, 30.2%
identity in 169 aa overlap and at the N-terminus, to
SW:RL25.ECOLI (EMBL:D13326), rplv, Escherichia coli 50S
ribosomal protein L25 (94 aa), fasta scores: E(): 3.5e-10,
40.0% identity in 90 aa overlap. Also similar to many
hypothetical proteins. Contains pfam match to entry
PF01386 Ribosomal_L25p, Ribosomal L25p family"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1094"
/protein_id="CAB84357.1"
/db_xref="GI:7379789"
/db_xref="SPTREMBL:Q9JUX7"
/translation="MTVEIQASVREAGTGASRRLRREGQIPGILYGEQGPVAVIAD
HKTFTYALKESEPTALIKLSLNGETKDVIYDFQMFHPRERVOHQHIDFQAVKADQVPR
IRVPLHIVNENSQAVKLOGGRVSLNTTVVEVVALPANIPAFULDCAEVVAGDILHL
SDILKEGESVSLKRNENLAVATVTGKR"
42006..42269
misc_feature 42006..42269
/genes="NMA1094"
/notes="Pfam match to entry PF01386 Ribosomal_L25p,
Ribosomal L25p family, score 128.60, E-value 1.1e-34"
42617..42652
stem_loop 42617..42652
/notes="Stem loop containing DNA uptake sequences: ttatat
gccgtctgaa ccgtgt ttcagacgcc atttc"
42622..42631
misc_feature 42622..42631
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(42638..42647)
misc_feature complement(42638..42647)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(42691..43860)
gene complement(42691..43860)
/genes="NMA1095"
```


/translation="MANGLDNOAFEDQVHFHTRADAPMQLAELSQKEMKETEGAALWFA
PVLGGRRALTCFTRHGLNOAISRGVGVSNKAINTWRNPTKITFSQSGRTRFTGA
DGVVLNSOGKALVTWTKRPR"
complement(29055. .29300)
/gene="NMA1079"
CDS
complement(29055. .29300)
/gene="NMA1079"
/note="NMA1079, len: 81 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1079"
/protein_id="CAB84342.1"
/db_xref="GI:7379774"
/db_xref="SPTREMBL:Q9JUY8"
/translation="MHEKPYKMMKNKIYLTAKAIGITAIISYIFLLNQESSGTERPPSY
FLMFNPLNLJLWFEENKTVMAAVTAIAAHFIFVALSD"
complement(29357. .29512)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 246.57"
/label=Correia
complement(29529. .29645)
/gene="NMA1080"
CDS
complement(29529. .29645)
/gene="NMA1080"
/note="NMA1080, len: 38 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1080"
/protein_id="CAB84343.1"
/db_xref="GI:7379775"
/db_xref="SPTREMBL:Q9JUY7"
/translation="MELKYCNREAKASVVTYIYFILISSKKEIQDYTCQ"
complement(29704. .29991)
/gene="NMA1081"
CDS
complement(29704. .29991)
/gene="NMA1081"
/note="NMA1081, possible membrane protein, len: 95 aa;
unknown, contains hydrophobic, possible membrane-spanning
regions, lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84344.1"
/db_xref="GI:7379776"
/db_xref="SPTREMBL:Q9JUY6"
/translation="MDDLIAAFGLCLSGKYSEKISRLEISLTAGSIFFAVFFIHILI
IFIVDGNKENLWNFITGIMKDLFLFSCVCLLFYLLFAETIRKTINKIK"
30052. .30114
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 78.71"
/label=Correia
complement(30298. .30669)
/gene="NMA1082"
CDS
complement(30298. .30669)
/gene="NMA1082"
/note="NMA1082, len: 123 aa; unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1082"
/protein_id="CAB84345.1"
/db_xref="GI:7379777"
/db_xref="SPTREMBL:Q9JUY5"
/translation="MRFYIDNKECDYIFFRKAVNISDDFIQSQFITPEGLITKSNPK
LENQYSKMYSKNCQYEMVTFLSDEMKNIELLSNDADIEFAYSDFVFLPEIETDK
RTLKSLNFYIDFGVKYIINKI"
30746. .30787
complement(30803. .31246)
/gene="NMA1083"
CDS
complement(30803. .31246)
/gene="NMA1083"

/note="NMA1083, possible periplasmic protein, len: 147 aa;
unknown, similar to parts of other proteins from Neisseria
meningitidis e.g. NMA1073, fasta scores; E(): 1.4e-17,
65.6% identity in 93 aa overlap, also to NMA1078. Contains
probable N-terminal signal sequence. Lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84346.1"
/db_xref="GI:7379778"
/db_xref="SPTREMBL:Q9JUY4"
/translation="NKKQITSAVMMLSMIAPAMANGLDNOTFENOVHFTQADAPMOLA
ELSQKEMKETEGAAPLVAIGILHAGRFLAORWVTORVAAQALSRGANVYARTSQAR
AVANQAWERQNVIRHQKEIHSNYSHFQNSQKIRGHAFYGNKHR"
complement(31252. .31256)
/gene="NMA1084"
CDS
complement(31289. .32158)
/gene="NMA1084"
/note="NMA1084, possible periplasmic protein, len: 289 aa;
shows weak similarity to TR:O53036 (EMBL:U68399), hmcD,
Haemophilus influenzae protein at haemocin immunity locus
(297 aa), fasta scores; E(): 3.7e-07, 24.6% identity in
289 aa overlap. Similar to part of NMA0171, fasta scores;
E(): 1.2e-09, 54.7% identity in 53 aa overlap. Contains
probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84347.1"
/db_xref="GI:7379779"
/db_xref="SPTREMBL:Q9JUY3"
/translation="MKRIFLPALPLLSAYADLPITIEDIMTDGKKWKLSTLYL
NSBNRAELAAPVYIQTGFAPIPTEIQENGSNTDMLVGLTGKRGYLTGNTDIYGS
SSYLWHEERKLDGNGKTRNKSDSLGSISHTFLKDDKNPALISLESTVYKSRNKA
SSGSKWLIGATTYKAIDPVVLSLTAAAYRINGSKTLSSNTKYKAGNYMLNPNISFAAN
DRISLTGGIOWLGKQPDRLDGKKESARNTSTVAHFAGFGFTKTTALNASARFNVSQ
SSSELKFGVQHTF"
complement(32163. .32167)
/misc_feature
32281. .32290
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
32322. .33245
/gene="NMA1085"
32322. .33245
/gene="NMA1085"
/note="NMA1085, len: 307 aa; similar to TR:O07166
(EMBL:296070) Mycobacterium tuberculosis hypothetical
protein (305 aa), fasta scores; E(): 0. 42.9% identity in
282 aa overlap and shows weak similarity to many,
including pseudouridine synthases. Contains pfam match to
entry PF00849 YABO, Hypothetical yabo/ycec/stfB family"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1085"
/protein_id="CAB84348.1"
/db_xref="GI:7379780"
/db_xref="SPTREMBL:Q9JUY2"
/translation="MKRKNPNPLILNGKPSYLVLPHEKQFYGLPLLLHFLCIRFPFVG
ADDMRRRLNSGFVSGDGAALDEHSLFEPGKVMFYRETRESEPRIPPEERILHIDE
HLIVVDKPHFLPVIPSGRFLRETLRLRLELOHNVEDITPLRLDKDTAGYMLL
SHNPATRGAYQPMFQNTKVTWKTYEALAPTRDLPYPLDVVSRVLRGKEFTTQEAEGE
PNAHTVLIENRGEFSLYRLTPHTGKGRHOLRVHMLGVLNLDALYVPPSEAGSED
YKPKLLAKKIAFAADPLSGSERVFCSGFL"
32613. .33095
/gene="NMA1085"
/note="pfam match to entry PF00849 YABO, Hypothetical
yabo/ycec/stfB family, score 131.20, E-value 1.9e-33"
33125. .33134
/gene="NMA1085"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

| | |
|------|--|
| gene | complement(25256.. .25456) /gene="NMA1070" |
| CDS | complement(25256.. .25456) /gene="NMA1070" /note="NMA1070, len: 66 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1070" /protein_id="CAB84333.1" /db_xref="GI:7379765" /db_xref="SPTREMBL:Q9JU25" /translation="MOLICADWTGIGSFMTKFEKTWSAQYRDMESISVRNWNLERTGA EVIYNGRRVYHNEAEAMASLR" 25460.. .25615 /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 271.26" /label=Correia complement(25571.. .25990) /gene="NMA1071" /gene="NMA1071" /note="NMA1071, len: 139 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1071" /protein_id="CAB84334.1" /db_xref="GI:7379766" /db_xref="SPTREMBL:Q9JU24" /translation="MSFGYLIATSQPCELLTKSGRTFSLIMDKMDLWIYFRCEGNI YTKNKTESCLTERGKWLKHIVEYNRGSIPIFYLLKRESEENPAEIVLAKSRNN KILTVKVRSLHFDLRNIYRIEMYSGLNLNQYVASP" complement(25997.. .26000) complement(26048.. .26413) /gene="NMA1072" complement(26048.. .26413) /gene="NMA1072" /note="NMA1072, len: 121 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1072" /protein_id="CAB84335.1" /db_xref="GI:7379767" /db_xref="SPTREMBL:Q9JU23" /translation="MEKPLTTPFFLFEDVSLDFQGLSELEKKIIEPQDLMDVYRAF DSVGNILNFRIVEKQKGFVWSTKIKTVVFDSDADSSDDLFLKCLQSSYKAYFETEP GLDKROLMTLQKCGFSC" complement(26395.. .26940) /gene="NMA1073" complement(26395.. .26940) /gene="NMA1073" /note="NMA1073, len: 181 aa; unknown, similar to parts others from Neisseria meningitidis e.g. NMA1083, fasta scores: E(): 3.3e-17, 65.6% identity in 93 aa overlap, also to NMA1078. Lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1073" /protein_id="CAB84336.1" /db_xref="GI:7379768" /db_xref="SPTREMBL:Q9JR36" /translation="MKKQITAAVMMLSIAPAMANGLDNQAFEDQVFHTRADAPMOLA ELSQKMKETEFALPAILGGAIGMTQHGFSYATTGRPASVRDVAIAGGIGATPG GVGAACKVVSFAKYGREIKIGNNMRIAPFGNRTGHPHFRVYHRRVTDNTGKTLPGQ GICRHPWESKTSRDKNRF" complement(26946.. .26950) complement(27103.. .27375) /gene="NMA1074" complement(27103.. .27375) /gene="NMA1074" |
| RBS | |
| gene | |
| CDS | |
| gene | complement(25256.. .25456) /gene="NMA1070" /note="NMA1070, len: 66 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="putative membrane protein" /protein_id="CAB84337.1" /db_xref="GI:7379769" /db_xref="SPTREMBL:Q9JRH1" /translation="MFAVIFFTGICLAWRIDPIKIKSKILARSILYIIGIINVIIS YVLIKNLIVSVDGGGKIVAVILSNLFFWTLMYVLVKRLSKKPS" complement(27383.. .27388) complement(27444.. .27614) /gene="NMA1075" complement(27444.. .27614) /gene="NMA1075" /note="NMA1075, len: 56 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1075" /protein_id="CAB84338.1" /db_xref="GI:7379770" /db_xref="SPTREMBL:Q9JU22" /translation="MGLGTCVCCIFGRNSGFEKIGRILINGMRGVAVTGAGGINY AGSTGKNTDIRR" 27644.. .27904 /gene="NMA1076" 27644.. .27904 /gene="NMA1076" /note="NMA1076, len: 86 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1076" /protein_id="CAB84339.1" /db_xref="GI:7379771" /db_xref="SPTREMBL:Q9JU21" /translation="MFLILTSLVQIKETDYTKDRHTENTYPLPCLNLTIDPIDQ TEKQSHSINQRKCYVKHNGFYLTVVYFNFSNLISEY" complement(27997.. .28151) /note="Correia element; hmms hit to HMM Correia (1 - 156), score: 260.70" /label=Correia complement(28251.. .28499) /gene="NMA1077" complement(28251.. .28499) /gene="NMA1077" /note="NMA1077, len: 82 aa; unknown, lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1077" /protein_id="CAB84340.1" /db_xref="GI:7379772" /db_xref="SPTREMBL:Q9JU20" /translation="MNKKEIVALTISKIKIDGGQYSLENFALSVLREKIKNVSSDE KLIIINDFLSGFLFDPETGKPVGETLKTIEQMIDFLV" complement(28545.. .28913) /gene="NMA1078" complement(28545.. .28913) /gene="NMA1078" /note="NMA1078, len: 122 aa; unknown, similar to parts others from Neisseria meningitidis e.g. NMA1073, fasta scores: E(): 8.4e-14, 64.6% identity in 79 aa overlap, also to NMA1083. Lies within a region of unusually low GC content" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA1078" /protein_id="CAB84341.1" /db_xref="GI:7379773" /db_xref="SPTREMBL:Q9JUY9" |

/note="NMA1063, probable GTP-binding protein, len: 485 aa; similar to many hypothetical GTP-binding proteins and to TR:O87407 (EMBL:AF058711) Neisseria gonorrhoeae essential protein (485 aa), fasta scores; E(): 0, 96.7% identity in 485 aa overlap. Contains two PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein"
/protein_id="CAB84327.1"
/db_xref="GI:7379759"
/db_xref="SPTREMBL:Q9JV01"
/translation="MKPTIALVGRPNVGKSTLFRNLRTTKDALVHDLPLGLTRDRHYGH
KVGSGPYLIVDTGPEPVVDSGILHMAKOTLOAVDEADVFLVDGRTGLTPDQKI
IADRLQSPRPVYLAVNKGEGNRAVLAEEFELAGLDPYVIGSAGHGQVYVLIEDL
ETFPPEKEEEAKHPYFAVIGRPNVGKSTLYNALIGEERVIAFDNAGTIRDSIHDF
ERGGFTIIDTAGVRRRGKVDAAVEKFSVIKQAVEANAVLVLDQOQDIADQDA
TIAGFALAGRALVAVNKGWDGISEERREQVRDINRKLYFLDFAKFHFISALKERGI
DGLFDSIQAYNAAMIKMPTPKITRVLQSAIERQOPRAGLVRPKMRYAHQGGMNPV
IVVHNSLHAISDSYTRYLTOTFRKAFNLQGTPLRIQYNVSENPNENADDKPKKKPLR
RVSLSNRIEGRGKEEKRFKKTKVSKQFQSK"
complement(20875..20898)
/gene="NMA1063"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(21394..21417)
/gene="NMA1063"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(21583..21592)
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(21595..22224)
/gene="NMA1064"
complement(21595..22224)
/gene="NMA1064"
/note="NMA1064, possible membrane protein, len: 209 aa; similar to SW:GC_NEIGO (EMBL:AF058711) Neisseria gonorrhoeae unknown protein (209 aa), fasta scores; E(): 0, 98.1% identity in 209 aa overlap and to bacterial hypothetical proteins. Contains hydrophobic, possible membrane-spanning region"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84328.1"
/db_xref="GI:7379760"
/db_xref="SPTREMBL:Q9JV00"
/translation="MAAHLEEQQELDNFKYFWKTGKWLFPAVLIILAALGYLYTVQV
RAASQNEAAVLANIVEKAQKPOSEINAEIAKIQSYPHSISAAQATLMAAAATEF
DAQRYDAEGHLKWLNSQKDSLIIQALAAQRLGVLLQOKKYDAALALDTPVEADFA
PLIMETKGDVYAAQGSQALKNYGOALEKMPQDSVGRELVQMKLDSLK"
complement(22225..23520)
/gene="hiss"
complement(22225..23520)
/gene="hiss"
/EC_number="6.1.1.21"
/note="NMA1065, hiss, probable histidyl-tRNA synthetase, len: 431 aa; similar to many e.g. SW:SVH_ECOLI (EMBL:M11843), hiss, Escherichia coli histidyl-tRNA synthetase (ec 6.1.1.21) (423 aa), fasta scores; E(): 0, 53.7% identity in 423 aa overlap. Contains Pfam match to entry PF00587 tRNA-synt_2b, tRNA synthetases class II (G, H, P and S) and PS00339 Aminoacyl-transfer RNA synthetases class-II signature 2"
/codon_start=1
/transl_table=11
/product="putative histidyl-tRNA synthetase"
/protein_id="CAB84329.1"
/db_xref="GI:7379761"
/db_xref="SPTREMBL:Q9JU29"
/translation="MAQKIQSVKGNMDDLPEQKDFKLTAAFWQAFEDTVGRWRTYTG
YQIQIRTPYVEQGLFVRSIGEETDVVGKEMTYFSDNSLTSAPGEGTASCLRAVE
HNFLYNSPKLMYGMFMRPRQKRYQFQHVGEALGFEQPDIDAEIAMSADLW
EKLIGREYLTLEINSIGNREERAHRAALVLEYLTRIYEDKLEDEDSKRRLKTNPLRLVDT

KNPDLQEIICNAAPRLVDYLGEASQNHYARFKAMLDGLGIQYIENPRLVGLDYINQTV
FEWTKQAGATVCGGVGDLIELGGKPAPSIGFAMGIERLLILLVSYGSLVYNA
APDYVAMHOGEGADLOVMKYAQALRAQGNVMQHSYGQSLKAQMKRADNSGARFALT
AQDELANGVTCLKDMGAHQDQTVAAADLTNTLQOMKNA"
complement(22234..22238)
/gene="hiss"
complement(22282..22337)
/gene="hiss"
/note="Pfam match to entry PF00587 tRNA-synt_2b, tRNA synthetases class II (G, H, P and S), score 203.10, E-value 4.3e-57"
complement(22552..22581)
/gene="hiss"
/note="PS00339 Aminoacyl-transfer RNA synthetases class-II signature 2"
complement(22895..22904)
/gene="hiss"
/note="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(23614..24270)
/gene="NMA1066"
complement(23614..24270)
/gene="NMA1066"
/note="NMA1066, possible periplasmic protein, len: 218 aa; shows weak similarity to part of TR:O53037 (EMBL:U68399), hmcC, Haemophilus influenzae protein implicated in haemocin production (198 aa), fasta scores; E(): 4.3e-12, 33.6% identity in 146 aa overlap. Also highly similar to other proteins from Neisseria meningitidis, NMA0173, fasta scores; E(): 0, 70.7% identity in 164 aa overlap and NMA0767, fasta scores; E(): 0, 88.3% identity in 128 aa overlap. Contains a probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84330.1"
/db_xref="GI:7379762"
/db_xref="SPTREMBL:Q9JU28"
/translation="MMKFYVLLACVYVLSVRLNAAPFNDNPVYVGKIKVQSHKE
RDFNIKQDLDFSCGAASVAILLNNFYQGLTEEEVLRKLDKEQMRASFEDMRRITMP
DIGFEAKGYALSFQALQKIPVIVLYKRYKDDHFSVLRGIDGNTVLLADPSSLGHVSM
SRAQFLDAWQTRGNNLAGKILAVVPKKAETISNKLFFTHHPKQTEFAVGQIRQARAE"
complement(24328..24519)
/gene="NMA1067"
complement(24328..24519)
/gene="NMA1067"
/note="NMA1067, len: 63 aa; unknown, lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1067"
/protein_id="CAB84331.1"
/db_xref="GI:7379763"
/db_xref="SPTREMBL:Q9JU27"
/translation="MLPTFSVSLLLGEYLEPEENGTKIAAETDSAWHFLGAACRTSI
NGKYVAGNRIYVFAKKAES"
complement(24552..25088)
/gene="NMA1068"
complement(24552..25088)
/gene="NMA1068"
/note="NMA1068, len: 178 aa; unknown, lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1068"
/protein_id="CAB84332.1"
/db_xref="GI:7379764"
/db_xref="SPTREMBL:Q9JU26"
/translation="MEKIIQSVRETNFNQARVIRAMOINISDEQVVIVRMSCYLK
KDEIFKNERYQSSDIKKSPFPPLNLPVPKPYFLYDLTNLQIYDLAWSCNDF
LMFEFILDRESIKYIDQDKLFKRETLTYSNFKSNQCKFFCGDFDAPYFKSGV
SQVADYNFVPDASLFFI"

/db_xref="GI:7379753"
/translation="NAVPPNLTGIFNRPPEQREKNKPHFVCTVSHLRRLKGDV
LLTAFARALACQPOLRLNIGSGQEQRLKQAAOLGITHAVTFGLGALQPEAVLDLMR
NSDAFLASRTETFGVYIEALSGPLVIATRCGGAESIVSDGNYLVPVDDDDALAD
ALIKMYHHSDFEPFARLENCLNEFGENAVIGRLIGIFRQAIATAEYGRKIPVKYSLGTLK
ITRRRSRQYK"
complement(15922. .16446)
/gene="NMA1057"
/note="Pfam match to entry PF00534 Glycos_transf_1,
Glycosyl transferases group 1, score 90.40, E-value
9.5e-25"
/transl_table=1
/product="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(16302. .16994)
/gene="NMA1058"
CDS
complement(16302. .16994)
/gene="NMA1058"
/note="NMA1058, len: 230 aa; unknown, shows weak
similarity to part of TR:O32272 (EMBL:Z99122), tuac,
Bacillus subtilis hypothetical protein (389 aa), fasta
scores: E(): 0.0033, 30.3% identity in 132 aa overlap. GC
frame and codon usage plots indicate a change of coding
frame at codon 189 (approx.), near a poly-A tract. This
suggests that NMA1058 may represent the original
N-terminal half of NMA1057, now frameshifted
has a Pfam PF00534 Glycos_transf_1, Glycosyl transferases
group 1 domain in its C-terminal half (absent from this
ORF, but present in NMA1057)"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1058"
/protein_id="CAB84322.1"
/db_xref="GI:7379754"
/db_xref="SPTREMBL:Q9JV05"
/translation="MWTGFFSKSGTGAERHQNRRACTDVPLLAERNKHPDRPYGF
AKYRQSLDIYAHGIVFFPRFIDIDCIRWVCAGLKAFKHITRENLPLDIIAHHRM
NYAGLQIKSOKYGIPIVVTHEHSTITRGLRPHQPMKEKAHAHSAALLAVSRHEFA
HVLQHKYGCWQYLLPPYWAEYSTDLNKEKKTINRISCSAPSTFAVSKDTMSYSLP
LPGWHNARCA"
17175. .17178
17186. .17890
/gene="NMA1059"
17186. .17890
/gene="NMA1059"
/note="NMA1059, possible periplasmic protein, len: 234 aa;
unknown, contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=1
/product="putative periplasmic protein"
/protein_id="CAB84323.1"
/db_xref="GI:7379755"
/db_xref="SPTREMBL:Q9JV04"
/translation="MLRSILAAISLAVSFPTAAEALNIVFESESAGVEAVQDTMSA
RFQVTAEGROKNVNAEFVKFNFTKNSFKTLVSRFSAMPRYQYTNGRIOG
WEAREAFKVEGRNFDALNRFIADVQADALEYTDHYSRERNRNVIOQSKDAVLRFK
ABEAKLAGVLCASGYKIVKLNLGHIGSHIAGGGAQAQMLRAMPMAASVNMEGADSA
PGVEEISVNGTVQF"
17905. .17949
/note="Stem loop containing DNA uptake sequences: caaat
gccgtctgaa accgcagcataaggg ttcacagcgc attta"
17910. .17919
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(17935. .17944)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(17954. .18520)
/gene="NMA1060"
CDS
complement(17954. .18520)
/gene="NMA1060"
/note="NMA1060, len: 188 aa; unknown, shows weak
similarity to SW:DCD_ECOL (EMBL:M90069), dcd, Escherichia

coli deoxycytidine triphosphate deaminase (EC 3.5.4.13)
(193 aa), fasta scores; E(): 0.00015, 28.5% identity in
151 aa overlap. Similar to many hypothetical deoxycytidine
triphosphate deaminases e.g. TR:Q928F1 (EMBL:AE001623),
dcd, Chlamydia pneumoniae putative deoxycytidine
triphosphate deaminase (190 aa), fasta scores; E(): 0,
67.6% identity in 188 aa overlap. Contains Pfam match to
entry PF00692 dUTPase, dUTPase"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1060"
/protein_id="CAB84324.1"
/db_xref="GI:7379756"
/db_xref="SPTREMBL:Q9JRE8"
/translation="MSIKSDKWIRMSSEFGMIDPFEPNQKEADGKRRIISYGTSSYG
YDIRCANEFKIFTINISTVDPKNFDKPNFTVEDDCIIPNSPALARTVEYFRIPR
NVLTVCLGKSTYARCGIIVNVTPEPEWEGYVTFESNTTLPKAIYAGEGVAQVLFF
ESDEICETSYKDRNGKYMGGTGVLPKA"
complement(17960. .18430)
/gene="NMA1060"
/note="Pfam match to entry PF00692 dUTPase, dUTPase, score
5.10, E-value 0.0011"
complement(18528. .18531)
18568. .18577
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(18586. .19002)
/gene="NMA1061"
complement(18586. .19002)
/gene="NMA1061"
/note="NMA1061, possible periplasmic protein, len: 138 aa;
unknown, contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=1
/product="putative periplasmic protein"
/protein_id="CAB84325.1"
/db_xref="GI:7379757"
/db_xref="SPTREMBL:Q9JV03"
/translation="MLKHLAFLFAMFALPAASAVLTSYQPGCTYEGDVKGDKGPA
KGTWRCDQGRNYTGSFKNGKFDGQGVYTVAAANREIFTEPFNSDKTKFRNMVLSGTFK
KGLAHRFTVSQNGETLFIKCMGMIKEVKLPKNK"
19041. .19050
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(19054. .19953)
/gene="rdgc"
complement(19054. .19953)
/gene="rdgc"
/note="NMA1062, rdgc, len: 299 aa; unknown, highly similar
to SW:RDGC_NEIGO (EMBL:AF058711), rdgc, Neisseria
gonorrhoeae protein which affects pilin antigenic
variation when mutated (299 aa), fasta scores; E(): 0,
99.3% identity in 299 aa overlap. Also similar to
SW:RDGC_ECOLI (EMBL:X76979), rdgc, Escherichia coli
recombination associated protein (303 aa), fasta scores;
E(): 0, 35.1% identity in 299 aa overlap"
/codon_start=1
/transl_table=1
/product="rdgc protein"
/protein_id="CAB84326.1"
/db_xref="GI:7379758"
/db_xref="SPTREMBL:Q9JV02"
/translation="MWFQISFYPLNKEKLPKPEADVLDKLAEEFTHCQGLDMFSEGF
TAPVSFSPFLVPADFTLRVALKKEKVLPGVIRDLKEKVAEIQNNEARNVGRKKK
QELKEQITDDLPRAFTRSRSTEAFTVTRHGYLLVNNAAKAENILTKREALGGLLE
ASLPNTKQSPSSLMTGWLLQGHCEGGFELSDCELKGTDIVPVKKVSKQDLTADDEV
QHVANGKVTQGLVWREQTAFILTDQFTLRKIQYLDVLOEAEESNGDDAAGLAFASQ
ILMAESVSIMLEELVSYLGGNQD"
complement(19960. .19963)
complement(19984. .21441)
/gene="NMA1063"
complement(19984. .21441)
/gene="NMA1063"

Escherichia coli single-stranded-DNA-specific exonuclease
(EC 3.1.1.-) (577 aa), fasta scores; E(): 0, 47.9%
identity in 584 aa overlap. Contains Pfam match to entry
PF01368 DHH, DHH family"
/codon_start=1
/transl_table=11
/product="putative single-stranded-DNA-specific
exonuclease"
/protein_id="CAB84317.1"
/db_xref="GI:7379749"
/db_xref="SPTREMBL:Q9JV09"
/translation="MSVKIQTRSVNTDVFNHLITAGADPLIARICASRGVQSPAELEDG
KLASLGYQTLTNCFAAARLADAVEQKELIIVADYDADGATACAVGLDGLAAMGAK
VHLVPPNRFHGYGLTPELAEIAAAGVDLLITVDNGIAGTAGARAQALGLDVIYVD
HHLPAETVPDCIIVPNOKGCPSPKSLAGVIFYVILMALRAELRRNRYFSGLKEP
NLGDLDLVALGTADVADVSLDHNRIILVSOGLKMRSEKMRPGITRALFEVARBWRKA
OPDMGFALPRINAGRLDDMSVGLACLLARDSDSEAQLAAQNLNINIERLEIQSM
LQDALNAFPETLPFGQMTLVAYDDPHQGVVGIIVASRLKDRFYRPTIVFAPADNGEVR
GSGRSIPNLHLRALDLVSKRHPDLILKFGGHAMAAGLSLEHNIIPAFQTFEEAVRE
MVCEDDLSTQFFITDGLSPACDITLEQAQNLARHWGOGFAPPSTDFEHVVRQOPLGA
EGKHKKVWLQKDGCEFEAFMWRCSEDIPEYIRTVIRPVANERNNLEQLYIDYWEAA
"
10939..12084
/gene="recJ"
/note="Pfam match to entry PF01368 DHH, DHH family, score
351.60, E-value 8.6e-102"
complement(11370..11379)
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
12686..12695
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
12739..14100
/gene="pcnB"
12739..14100
/gene="pcnB"
/EC_number="2.7.7.19"
/note="NMA1053, pcnB, probable poly(A) polymerase, len:
453 aa; similar to many e.g. SW:PCNB_ECOLI (EMBL:M20574),
pcnB, Escherichia coli poly(A) polymerase (EC 2.7.7.19)
(472 aa), fasta scores; E(): 0, 41.3% identity in 441 aa
overlap"
/codon_start=1
/transl_table=11
/product="putative poly(A) polymerase"
/protein_id="CAB84318.1"
/db_xref="GI:7379750"
/db_xref="SPTREMBL:Q9JV08"
/translation="MLKWLKMLPSGRSSKKAESKTVIPAERHNTRAEMLSFAAENV
IRLKGAGFOAYVVGAIIRDLLIGPKDFDVAIDAMPEOVHKLFRSRIVGRRFOIV
HVNGAELIEVTFRGKAKVHONAGRIKMDNTYGSIEEDAMRRDTCNALYIDPEKE
EILDFFNGIADVAARLYMIGDAERYQEDPVILRAIRLSGLGFELSEETAPPIAE
SICRLKHEPVARLEIMKLLFSGHARECLRLNGEDIPDDIHLPLNALRVSDGIAGK
MYLAKNDELRLRADKSVSGFVLAALMWPPELERHWSNLQQLKLPVPALSDAINTM
RETVERGWPQPSFATMREIMWFQFENRKGARPHKLPFAQARFRAAYDFLLLRAET
GNADRALAEWTFATQASQQRTEMTKNEAAARHEKNEGAKRRRRRRKPKPKVVG
DWP"
12875..12884
/gene="pcnB"
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
13400..13409
/gene="pcnB"
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
complement(13545..13554)
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
complement(13955..13964)
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
14125..14277
/note="REP 2; hmms hit to HMM REP 2 (1 - 128), score:

105.80"
14262..14265
14270..14596
/gene="NMA1054"
14270..14596
/gene="NMA1054"
/note="NMA1054, possible periplasmic protein, len: 108 aa;
unknown, contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB84319.1"
/db_xref="GI:7379751"
/db_xref="SPTREMBL:Q9JV07"
/translation="MMKKGILGILTACAAMPADFADRIGLDEARLAQLERHVRVAVLESG
NTVKIDLFGSNSTMVYCVTPFQKTFEASDRNEGVARQKVRQACNRRTSAMFCDEAI
RCKRFD"
complement(14682..14691)
/note="Core DNA uptake sequence: gcggtctgaa"
/label=DUS
complement(14701..15654)
/gene="NMA1056"
complement(14701..15654)
/gene="NMA1056"
/note="NMA1056, len: 317 aa; unknown, similar to many
bacterial proteins of the phoH family e.g. SW:PHOL_ECOLI
(EMBL:AE000170), ybeZ, Escherichia coli hypothetical
protein (359 aa), fasta scores; E(): 0, 58.1% identity in
320 aa overlap. Contains P500017 ATP/GTP-binding site
motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1056"
/protein_id="CAB84320.1"
/db_xref="GI:7379752"
/db_xref="SPTREMBL:Q9JV06"
/translation="MTHVHLHEETDNLALQRCGSFDNNLLAKALDIHSRRPE
HFTFNAGFAHAGKRALKLELTATQDLNDGDIRLAAVEAQTEDAGHOENKDHAYVF
RTFRSGTGTTPRONGYIRALLNHDIVFGLPGAGTCKTYLAAVAAVDAMEKHQVERII
LVPRAVEAKGLFLPGDLTKQKDPVLYRPLDYDALYDMGFDRTVKLIEKGLIEIPIA
DAREKTLNGAYIILDEAQTNPQOMKMFUTRIGFGAKAVITGDTSQIDLPNKIKSGLK
DAREKILGVEGLYHFTFTGEDVVRHPLVQKIVEAYESAHD"
complement(15235..15258)
/gene="NMA1056"
/note="P500017 ATP/GTP-binding site motif A (P-loop)"
complement(15694..15849)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 272.96"
/label=Correia
complement(15793..16485)
/gene="NMA1057"
complement(15793..16485)
/gene="NMA1057"
/note="NMA1057, possible glycosyl transferase, len: 230
aa; shows weak similarity to many putative glycosyl
transferases e.g. to the C-terminal half of TR:AAD35706
(EMBL:AE001736), TM0622, Thermotoga maritima putative
lipopolysaccharide biosynthesis protein (388 aa), fasta
scores; E(): 1.5e-15, 31.9% identity in 226 aa overlap and
SW:WCAL_SALTY (EMBL:X56793), wcal, Salmonella typhimurium
putative colanic acid biosynthesis glycosyl transferase
(406 aa), blastp score: Expect = 1.0e-08. Contains Pfam
match to entry PF00534 Glycos_transf_1, Glycosyl
transferases group 1. GC frame and codon usage plots
indicate a change of coding frame at codon 21 (approx.),
near a poly-A tract. This suggests that NMA1058 may
represent the original N-terminal half of this CDS, now
frameshifted. Alternatively, they may both be unrelated
pseudogenes"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase"
/protein_id="CAB84321.1"


```

/misc_feature
/feature="clpA"
/feature="PS00870 Chaperonins clpA/B signature 1"
complement(6173..6196)
/misc_feature
/feature="clpA"
/feature="PS00017 ATP/GTP-binding site motif A (P-loop)"
6374..6383
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(6842..7153)
/feature="NMA1046"
complement(6842..7153)
/feature="NMA1046"
/feature="NMA1046, len: 103 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YLJA_ECOLI (EMBL:AE000190),
yjiA, Escherichia coli hypothetical protein (103 aa),
fasta scores: E(): 7.5e-14, 44.6% identity in 83 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1046"
/protein_id="CAB84312.1"
/db_xref="GI:7379744"
/db_xref="SPTREMBL:Q9JV14"
/translation="MNHPTDHQSDLLSDINTQPKRYGVFLNDDYTTMEFVVEL
TEVFLAQEAQAVAMLLLVHHEGKGLCSYTRDIAQTKQHVMERATEGHPLKACIVEE
V"
complement(6847..6851)
/feature="NMA1046"
complement(7171..7180)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
7399..7404
7411..7614
/feature="cspA"
7411..7614
/feature="cspA"
/feature="NMA1047, cspA, probable transcriptional regulator,
len: 67 aa; similar to many cold-shock proteins e.g.
SW:CSPA_ECOLI (EMBL:M30139), cspA, Escherichia coli cold
shock protein (69 aa), fasta scores: E(): 1e-15, 64.5%
identity in 62 aa overlap. Contains PS00352 'Cold-shock'
DNA-binding domain signature"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB84313.1"
/db_xref="GI:7379745"
/db_xref="SPTREMBL:Q9JV13"
/translation="MATGIVKWFNDKAGFGFITPDESGEDLFAHFSAINWDGFKTLKE
GQRVSFDVTGPGKQAANIQA"
7453..7512
/feature="cspA"
/feature="PS00352 'Cold-shock' DNA-binding domain signature"
7453..7605
/feature="cspA"
/feature="cspA"
/feature="Pfam match to entry PF00313 CSD, 'Cold-shock'
DNA-binding domain, score 107.50, E-value 1.6e-28"
complement(7889..7898)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(7912..9243)
/feature="NMA1048"
complement(7912..9243)
/feature="NMA1048, len: 443 aa; unknown, similar to many e.g.
SW:PMBA_ECOLI (EMBL:X54152), pmBA, Escherichia coli
protein possibly involved in the control of DNA gyrase
(450 aa), fasta scores: E(): 0, 50.4% identity in 425 aa
overlap. Contains Pfam match to entry PF01523 pmBA_TlIdD,
putative modulator of DNA gyrase"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1048"

```

```

/protein_id="CAB84314.1"
/db_xref="GI:7379746"
/db_xref="SPTREMBL:Q9JV12"
/translation="MLFNHTASELLDLCRTLDLAKATGATAAFADEFSIGOSVSVR
LGLEIOEFQODKSLDITVYVGRKGRAGTADSEQALRTVRAADIAHRHTADGCA
GLADPELMAOHIGDPDLYHEWDLDTAAAGLADKAGTADSEQALRTVRAADIAHRHTADGCA
YQVYGNTRYGFAAHRGTHHSISCSVVAADENGMDQRYWYDSACRHDPDMSPTIGQT
AARTTLRLGRSIPHTGTVPLFDITVSGGLIGHVLGALSGGALYRQSSFLIDISIGKK
VLPDFLNRPEPHIPRSRSSFDAEGVATAPRFVQNGIVEGVFLSSYSARKLGMOT
TGNAGGAHNLNLNTHETOSDLLKMGCTGLLVTLMGOGANTITGDSRCAAGFWVEN
GVTAIPVHEITVAGRIQDMYRDIVGVADDAALRRSSNKISILLIAGMTVAGS"
complement(8284..9165)
/feature="NMA1048"
/feature="Pfam match to entry PF01523 PmbA_TlIdD, putative
modulator of DNA gyrase, score 182.80, E-value 5.6e-51"
9334..9909
/feature="NMA1049"
9334..9909
/feature="NMA1049"
/feature="NMA1049, len: 191 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YJGA_HAEIN (EMBL:U32794),
Hil151, Haemophilus influenzae hypothetical protein (178
aa), fasta scores: E(): 9.4e-19, 36.4% identity in 176 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1049"
/protein_id="CAB84315.1"
/db_xref="GI:7379747"
/db_xref="SPTREMBL:Q9JV11"
/translation="MPSERGIILLCAIGKNSRTKMFQEDEWISKTQMKKQNNGLQD
LGMELTKLNDTLTKIGLDELDYEAVVYKITSNGALKRQAFIGRLMRDTPAPIE
AFLAKLRGDDAHNAFLQRFVEQARVRLADDGALTQFMSDFHADAGKLTILRNTKK
EOBNKPKPNFRALFQELKTVMSQSGTGGA"
9336..9345
/feature="NMA1049"
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
9956..10669
/feature="NMA1051"
9956..10669
/feature="NMA1051"
/feature="NMA1051, possible membrane protein, len: 237 aa;
unknown, contains probable N-terminal signal sequence and
hydrophobic, possible membrane-spanning region"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84316.1"
/db_xref="GI:7379748"
/db_xref="SPTREMBL:Q9JV10"
/translation="MLFRKTTAAVLAATLMLNGCTVMMGMNSPFSETTARKHVDKDO
IRAFGVVAEDNAOLEKESGLVMGKGYFVYVNPEDSAKLTGLKAGLDKQFQVYENPR
FAYQALPVKLESPASONSFTGCLYDRDPADIAKLKOLEFAVELDNRTIYTRCV
SANGKIYATPQKLNADYHFQESVPAADIYTVTKKHTDKSKLFENIATPTTLLDAVG
AVLALPVAALIAATNSSDK"
10573..10707
/feature="Stem loop containing DNA uptake sequences: gcaat
gccgtctgaa aagcc ttacagcgc atttc"
10678..10687
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(10693..10702)
/feature="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
10738..12438
/feature="recJ"
10738..12438
/feature="recJ"
/EC_number="3.1.-.-"
/feature="NMA1052, recJ, probable
single-stranded-DNA-specific exonuclease, len: 566 aa;
similar to many e.g. SW:RECJ_ECOLI (EMBL:M54884), recJ,

```

Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES

```
source
1..331801
  /organism="Neisseria meningitidis Z2491"
  /strain="Z2491"
  /db_xref="taxon:122587"
  /note="serogroup: A"
  117..1333
    /gene="NMA1040"
  117..1333
    /gene="NMA1040"
  /note="NMA1040"
  /note="NMA1040, possible type I restriction-modification
system specificity protein, pseudogene, len: 1217 bp;
N-terminus shows weak similarity to the C-terminal half of
TR:Q50359 (EMBL:L25415), hsdS1B, Mycoplasma pulmonis
restriction-modification enzyme subunit S1B (336 aa),
fasta scores; E(): 2.3e-06, 26.2% identity in 141 aa
overlap and to many hypothetical restriction-modification
subunits. Also similar to NMA1041, fasta scores; E():
4.8e-10, 37.0% identity in 100 aa overlap. C-terminus
similar to part of SW-T1S1_ECOLI (EMBL:X13145), hsdS,
Escherichia coli type I restriction enzyme EcoR124II
specificity protein (410 aa), fasta scores; E(): 1.9e-14,
39.6% identity in 182 aa overlap. Similar to NMA1040, E():
2.9e-09, 38.0% identity in 100 aa overlap. Contains Pfam
match to entry PF01420 Methylase_S, Type I restriction
modification DNA specificity domain. Contains a G(8) tract
which would allow translation as an intact CDS, if
variable. Lies within a region of unusually low GC
content"
  /codon_start=1
  /pseudo
  /transl_table=11
  /product="pseudogene (putative type I
restriction-modification system specificity protein)"
  complement(157..166)
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  171..620
    /gene="NMA1040"
  /note="Pfam match to entry PF01420 Methylase_S, Type I
restriction modification DNA specificity domain, score
105.50, E-value 1.1e-27"
  699..706
    /note="(g)8"
  726..735
    /gene="NMA1040"
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  740..1231
    /gene="NMA1040"
  /note="Pfam match to entry PF01420 Methylase_S, Type I
restriction modification DNA specificity domain, score
158.70, E-value 1e-43"
  1405..1408
  1415..14509
  1415..14509
    /gene="NMA1042"
  /note="NMA1042, pseudogene, probable type I
restriction-modification system restriction protein, len:
291 aa; similar to parts of many e.g. SW:TIR1_ECOLI
(EMBL:X13145), hsdR, Escherichia coli type I restriction
enzyme EcoR124II R protein (EC 3.1.21.3) (1033 aa), fasta
scores; E(): 0, 73.9% identity in 291 aa overlap (frame
1), followed by a stop codon, E(): 0, 92.9% identity in
127 aa overlap (frame 1), followed by a frameshift near a
poly-A tract, E(): 0, 72.5% identity in 604 aa overlap
(frame 3)"
  /codon_start=1
  /pseudo
```

```
misc_feature
  /transl_table=11
  /product="type I restriction-modification system
restriction protein (pseudogene)"
  complement(1642..1651)
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  1687..1696
    /gene="NMA1042"
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  1761..1770
    /gene="NMA1042"
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  3906..3915
    /gene="NMA1042"
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  4514..4545
    /note="Stem loop containing DNA uptake sequences: aaaa
gccgtctgaa at ttcagacgc atttt"
  4519..4528
    /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  complement(4531..4540)
  /note="Core DNA uptake sequence: gccgtctgaa"
  /label=DUS
  complement(4559..6838)
  /gene="clpA"
  /note="clpA"
  complement(4559..6838)
  /gene="clpA"
  /note="NMA1045, clpA, probable ATP-dependent protease
ATP-binding protein, len: 759 aa; similar to many e.g.
SW:CLPA_ECOLI (EMBL:M31045), clpA, Escherichia coli
ATP-dependent Clp protease ATP-binding subunit (758 aa),
fasta scores; E(): 0, 56.2% identity in 762 aa overlap.
Similar to NMA1683, fasta scores; E(): 0, 37.2% identity
in 844 aa overlap. Contains Pfam match to entry PF00495
clpA_B, Chaperonin clpA/B, PS00870 Chaperonins clpA/B
signature 1, PS00871 Chaperonins clpA/B signature 2 and
two PS00017 ATP/GTP-binding site motif A (P-loop)"
  /codon_start=1
  /transl_table=11
  /product="putative ATP-dependent protease ATP-binding
protein"
  /protein_id="CAB84311.1"
  /db_xref="GI:7379743"
  /db_xref="SPTREMBL:Q9JV15"
  /translation="MLSPELEHILQLLYREARNARYEFISLEHLLLVLFEEAAPPV
LKLCGADLKVLSEQLAASVAENTPOIPDHLDTVETPTLGFQVIRAMVHTQSAGK
AAVEPDLILVAMSSSESHAVYFLKLOSTRFEVLRCIAHGSPDDEGNDSDGLGREG
EEAEQKTSLSDYTYNLAENEKAGRIDPLIGRKHMERLVQILRRRNPNLLVGEAG
VGKTAALAEGLAHQIVKGDIPOALKDAEYALDMSGLLAGTKYRGDFEARKVSKVLE
KIPHAILEIDETHTIGAGSTSGGTMDSNLKPAKALGALRCIGATTYDEYRTIFDK
DHALSRRFQKIDVPEPTVQILRGLKPMFEGHQVRYTQGLAEAAELSARYINE
RFLPKAIDVMDGAQORILPKSKQKVIKQAJETVIAKVARIKPTVSHDDQVL
QFLGDLKNVYQGENAIYALVAAYKMSRSLGLPDKPIGSLFSGPTGVGKTEYAKQ
LAYSMGVPLQRFQDMSEYMERHVSRLIGAPPVGVGFQGGLLTEAVNKPQHPCHVLLDE
IEKHPDIFNVLQVMDAGKLTNDNGSADFNIVLIMTNAGASLSRSLSGFTAKR
ERGDEMAYINKLFTPEFRNLDAIIPFAPLSEPIIVKVVDFKLLQLEHRLDKKVEAE
FTPALRYLAERGFDPQMGARPMNRLIQEKIRKPLADELLFGKLVGGFVRIDWDAAK
EEAVLKFKKSKVKPETVY"
  complement(4637..5578)
  /gene="clpA"
  /note="Pfam match to entry PF00495 clpA_B, Chaperonin
clpA/B, score 583.70, E-value 1.1e-171"
  complement(5228..5284)
  /gene="clpA"
  /note="PS00871 Chaperonins clpA/B signature 2"
  complement(5330..5353)
  /gene="clpA"
  /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
  complement(5894..5932)
  misc_feature
```

36104 CCGAAGACATGTGTGAAGAGGCATTTTCCCGCTTGTCTCGGGCTCCGGTCCGGGACATCG 36154 |-----| 36163
-----|
36164 ACACAGAGCGTGGCTGATGATCGTGTGTCACACAGCGGCTGTGCCTGGATCATCAAGT 36168
|-----| |-----| |-----| |-----| |-----|
36224 CCGGTGACACGCCGGGAGCGCCCGCAGGACATCGCCCGCATGGCACGACGAGTGCACGCA 36256 |-----| 36277
36225 36242
36284 GCGTGTATCGTTGACCGGGCTGACCGGGTGACTCTCGACACGAGGTCCGGCTGGCTT 36322
|-----| |-----| |-----| |-----| |-----|
36344 TCCTGATCATGTCTGAGGCGCTTCGGCCCGCGAGCGGTGCTTCTGTTGCTGCACGAGA 36359
|-----| |-----| |-----| |-----| |-----|
36404 TCCTTTGGGTGCCCTACACAGCAATGCGCCACGACGATTGGCAGCCAGGCTCCACATGCC 36427 36435
36464 GGCAGTGGCTCATCGGGCCCGTCGA |-----| |-----| |-----| |-----| |-----|
36491 AGATCAACGAATCGGCGCATTTGGGCCAGCGTGGAGCCAGCCOAGCATCGCGTCGTCAACA 36541 |-----|
36551 GAGCTTTCATCGAAGCCTGCTCCACAGGAGACCTGGACACCCCTGCTCGAGGTGCTGGATC |-----| |-----| |-----| |-----| |-----|
36611 CGGGTGTCGCGCGGAGATCGACGCCCGCAAGGCGTTGCTGCTGCGTGGCGCGGATCGGG 36616 |-----| |-----| |-----| |-----| |-----|
36616 36631 36643 36649
36671 TTGGCCCGACCATCTGCGCCACTGGAGTCACCC |-----| |-----| |-----| |-----| |-----|
...
36718 GTAGCCACGCGGTGTGCGGTCAACCGGGGTGCTGGCTTTGTCAACCGAGCGCTTGCC 36768 |-----| |-----| |-----| |-----| |-----|
36778 GCGGTGTTGGCCCTGCTGATCGAGCCGCGCAAGATCACAAAAATCC |-----| |-----| |-----| |-----| |-----|
...
36895 GGTATGACCATGAATCACTTGCGCGCTTGACCGGCGAGTGTTGTCATCGTCGCGG 36945 |-----| |-----| |-----| |-----| |-----|
36955 TGGCCTGGCAGCCCTACCTGCTGAGCACCATCATCAGGCGGCATCGCGGTTACCGATATC |-----| |-----| |-----| |-----| |-----|
37015 GCGGACGGCGCGCGGTGCTGTTCTTCACTGCGCAGCTGGAGCTTTGTCTGGCGTGAC 37015 |-----| |-----| |-----| |-----| |-----|
37075 GTGTTGTTGCGTCTAGCAACGATTTTCCGGTGTGTTGCCATCGACGCGACCGGTTGTGGG 37093 |-----| |-----| |-----| |-----| |-----|
37135 CTCACGACCGGCTCTCAACCCCGCAACACTTGGCCAGCGCGGCGATGCAATCACCTC 37138 |-----| |-----| |-----| |-----| |-----|
...
37232 TAGCCACAGACCTGGGCGGCCCGCGGCTTCTCTGGCGCGCCCGTCCGCGCGACCGG 37282 |-----| |-----| |-----| |-----| |-----|

37292 TCGCGGCACTGCGCGGGTCAACTGCTTCGCATGGCGGCCACGGGTCCGCTGTTCGGGG 37322 |-----| |-----| |-----| |-----| |-----|
37352 GCATGCTCGCGCGGATGGCGCGCCCGCTGCTGCTGAACCTGGACGCGGCGCATCAATGCCG 37362 37372
37412 TTGCCCGCGGACGCTCGACGCGGTTTCGGGGCGCGTTCGGCACTGGAGCCGCGAGACCGGG 37422 |-----| |-----| |-----| |-----| |-----|
37472 CGGCTT 37472
...
37511 CGTGGCATCCCTACTTCCCGGATCGCGCGCTGCGCCATGCCCTCTATACCGAGCTCGAGG 37561 |-----| |-----| |-----| |-----| |-----|
37567
-| 37572
37571 CCGCGTTGGGGGGGTCTGGCCGATCGGCCACTGCTGACCATCTTCGGTCACTTCAACG 37627 |-----| |-----| |-----| |-----| |-----|
37631 ATCCGCTGGGTTTCAGCGCGCTGGAAGAGTTGTTCCGACGCGCACGCCAACTTGCAGG 37632
37691 TCCGCGGGGCAACCACTTCCCATGTGTGACGACCCAGACCTGGTGGCGGGGCACTCA 37720 |-----| |-----| |-----| |-----| |-----|
37751 CGTCTTTCTGTCACAGGTCACGCTG 37751
...

2872 matches found in sequence:
nma422491 : TOIG of: nma422491 check: 8990 from: 1 to: 331801
(from "mycobacter_ge.seq")
TOIG of: nma422491 check: 8990 from: 1 to: 331801
LOCUS NMA422491 331801 bp DNA BCT 04-DEC-2000
DEFINITION Neisseria meningitidis serogroup A strain 22491 complete genome;
segment 4/7.
ACCESSION AL162755 AL157959
VERSION AL162755.2 GI:7379742
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis 22491.
Neisseria meningitidis 22491
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
AUTHORS
1 (bases 1 to 331801)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skellton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491
Nature 404 (6777), 502-506 (2000)
20222556
REFERENCE
MEDLINE
2 (bases 1 to 331801)
Parkhill,J.
AUTHORS
TITLE
Direct Submission
Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
JOURNAL
COMMENT

```
33801 TCAGTGGCCGGGGGAATG 33764
...
33837 GCCCGCTGACGGTTTCGTCTACGGATCACGCCGTTCAACTTCACCTTCGATCGCGGCA 33887
33897 ATCTGCCGACGCCCGCGCTCTGATGGGCAACACCGTGATCTGGAAGCCGTCGATCACCC
33957 AGCGCTGGCGGCCTATCTGACCATGCAACTGCTCGAGGCGCGGGTTGCGCGCGGGG 33958
34017 TGATCAACCTGGTCACCTGGCGACGGATTTCGCGGTTTCGGATGTGGCACTGGCCGATCCAC 34034
34077 GGCTGGCCGGCAT
...
34253 ATTCGCGAGCATTCGATTACACGGCCAGAAGTCTCGGCGGTCTCGCGAGCGTTTATC 34303
34313 GCGCATTCGGTGTGGCAGCGATGGCGGATGAGTTGCTGGCCAAAGCCGCGAGCTGGC 34337
34373 TAGGTGACATCACCGACCTCTCCAACCTACGGTGGTGGCTGATCGACACGGCGCCTTC
34433 GTCAAGAACGTTCGACGCGCATCGAACGGGCCAAGGCGCGGCGGTACCGGTGCGCGTC 34484
34493 GCGGCGGAATACGACGACGAGCAAGGCTATTTCGTGCGCCCGCACGGTGTTCCTCTCCGAC 34505
34553 GACCGACCGAGAGTCGTTTGTCTCATCGAGTACTCGGTCCGCTGCTGTCGGTGCATGTC 34567
34613 TACCCCGACGAGCGCTACGACGATCCTCGACGTCATCGACACCGGATCCCGCTACGCG 34623
34673 CTGACCGCGCGGTCTATCGCGAGCACCGGCGAGGCCGTGCTGACCGGCGTGGATCGGGTG 34688
34733 CGGTTCCGCGGGGAACTTCTATGTCAACGACAAAGCCGACGGGGGGGTGGTGGGCGT 34760
34793 CAGCGGTTCCGGGTGCACGCGGATCGGGCACCAACGACAGAGCGGTTCCCGCTTGAC 34826
34853 CTGCTCGGGTGGAGTTCGGCGCGCAGCATCAAGGAGAGCTTTCGTGCGGCGCACCGACAC 34864
34913 ATCTACCGGCACATGGCGGTGCGACTGATGGC 34888
...
```

```
31440 ATTACCGTTTCGGCGATCTCCGATAACCCAAAGCGGGCAGGGGATGTTGGCGATAGCG 31490 31496
-| |-----|
31500 ATGCTGATCATAAACGTTCTGCAATGCTGTGCATGTGCTGAAACAGGTTGACGCAGAGTCG 31511
|-----| |-----| |-----|
31560 AAGTCGGGTAGCGAGGGGCGCCGTGAGGGCGTCAACGGTCAGTTGCTAAAGCCGTGCGT 31577 31589
|-----| |-----| |-----|
31620 TCCATGGCCCGCAGCCCCCAGCGAAAGAGCAGCCGACATCCGGATCGCCAGCGAGGTC 31664
|-----|
31680 GACAAACAGCTGCTCGATCCGCTTATCCGGTAGCGAACGGTGTGGGATGCACCTTGCAAT 31700
|-----| |-----| |-----|
31740 GACCGTGGCGGGCGCGCGATGTCGCCGAAGCATCCAGGTAGGCACGAGGGTCTGAGCC 31751 31760
|-----| |-----| |-----|
31800 AGCACCGGGTCTCTGGGGCCCGAGTCACTATCCGAGGATCGACAGCCGCTGTCGCGTG 31814
|-----| |-----| |-----|
31860 CCGACCAAGGTGACGATTTCTGTCAGCAGAACGGTGGTGCCTCGGGCAGCGATGTC 31871 31911
|-----| |-----| |-----|
31920 ACCTGCCCAAGATCGGGTGGCGCTCGGCACTCTGAGTACCCGATCCACCTCAGCGGT 31939
|-----|
31980 GCCGGGTGACTTCG 31993
...
32103 GTGGTTTGGGCACGACACATAGATCCGTGAGCCGTGGCGGAACCTGAGCGTCTGTTG 32153
|-----| |-----| |-----|
32163 CGAAAGCGCTGGCGCTCAATGCCATGACGTCAACAGCCGAAACATGGCGGACTGCGGTA 32168 32174 32189
|-----| |-----| |-----|
32223 TCGCGGTTTTCCGGGTGTCGAACCCGATCAGCGTTGCGTTGCCCTCGGGCGGACGCCG 32253 32276
|-----| |-----| |-----|
32283 AGTTACGGCGGATGCTGATACGTCGACGGGTGCTGTGTTGCGTTACGCTCGGCCAGG 32291
|-----|
32343 CCCAGTAGTTGCTGTATACCCCGACGGGTGCGTATTGGGTGGGTGCGCCAGTCGCGACATG 32385
|-----| |-----| |-----|
32403 ATCCGGGGCCAGCAGCCGAGCACCCCGCAACATCTCCTCGCATCGTGGGCCAACGGC 32447
|-----|
32463 TGGAGCCTTGTGTGACCCAGATCTGTGCCGGGGAACACCG 32477
|-----|
... |-----|
```


26457
26492 TCGCGCGCCCAAGCTTATCTAGCGCCAGCTTGATTGGTCTGGTTCATTTGGCCAGCT 26515
26552 GCGGAGCCTGGCTCACTTCAACTAGGACGACCGCAACAAANTTCCCGCTTCGGATCCGA 26578
26612 GTTCGGTTGGGTACGGCGCAAT 26578
...
26711 GCGGCCTTGGCCGACCTGGAGCAGCTGGCGCAACGTGTGAGCCAGATCCAGGCGGTTCGCCA 26761
26771 TGGTTCCGGGTGTGACCGGGCCAAACGGGGAACCCCTTGAACAGCGCCGGCGGACATACC 26820
26831 AAGCGGGCCAAAGTTGGCAACCGGTGGCGCGCGTCCGGAATGATCGATGAGCGCACCG 26890
...
26891 GCGACCTCAATCGGCTGGCATCGGCTGCCAACCTGTTGGCCGACAATCTCGGTGACGTTTC 26944
26951 GCGGTCAAGTCAGCGCGGCGCGTTCGGGTGTCCGAGCGTTGTGACGCGCTTCGACGCCCTCGCTTACA 26995
27011 TCCAGAACCAAGTTCGGTGGCAACAAACATTTCAACGAAATCGACAAGCTGCAAGGCTTG 27055
27071 TCAGCAATATCCACGCGCTCGGTGACCGTCTGAGGTAAACTTTGACGGTATCCGCCAACA 27094
27131 GTTTCGATTGGCTTGACTCTGTCGCGCGCTTTGGATACCAAGCCGCTCTGTGACAGCA 27154
27191 ACCCTATGTGTGGCAACGC
...
27278 CGGCTCAGCTCCAGTCCACGCGGTACCGGAGACCGGTGTCGGCGGTGGTGAACGATCTGG 27328
27338 GCGGATCGCTGAATTCGGTAGTCCGCTCGCTGAAATCACTGGGGTTGGACAATCCGACG 27342
27398 GCGCGCGCGCGCTGATCAGCATGCAAAATGGAGCTAACGACCTCGCCAGCGCGGCTC 27399
27458 GTCAGTCCGAGACGCGGTCCAGATGCTGTCGACCAAGCAAGAACATGGGCATCGGGC 27460
27518 TGAACGAGCGGTGAGCGCTTCTGATGGCGATGGGCAACGATGGTTCGCAACCGCTCGATGG 27525
...
27543

27578 CCGGTTTCAATGTCCCGCCGCAAGTGTGAAGTCCGAGGAGTTCAAAAAAGTCGCCGAGG 27628
27638 CGTTTCAATCTCCAGACGGGCATACCGTGGGTACTTTCATTTCAGACCGACCTCA 27641
...
27803 CCGTCATGCTGAGGGACATCCGCCGATTACTACGAGCGGATATCGGCGTCACTGCTCGCTG 27853
27863 TGACCGTCTGCTGGTGTGATCCTGATCCTCATGCGACTGCTGCGTGCATAGTGCGCCGC 27915
27923 TGTACCTGGTGGTTCGGTGGTCACTCGTACATGTTCGGCGATCCGGCTTGGTGTGGTG 27960
27983 TGTTCAGGTGTTCTTGGGGCAGGAATTGCACTGGAGTGTGCCCGGCTACGCTTGTGTGG 28032
28043 TGCTGTGTCGCGCTGGGTGCGGACTACAACATGCTGTGGCGTTCGGGACGAGT 28080
28103 CCGCATTTGGAGTCCGTTCCAGCGTGATTCGCA
...
28175 TGATATTTGGCGCTTCGATGTCCGGCCTGCTGTCTCCAGCATCGGAACGCTCTCTCAAG 28225
28235 GCGGCTTCATATTGGGGTCGGATCCTGTAGACACGTTCTCGTGGT
...
28520 GGATCCAGGCTCGCACCGGATCCACGAGACCGCGCGGTGAACCAATTGCTGTTCCGT 28570
28580 GCATAGCGCGGTCTGATCTGTGGTTGCAGCACACTGCTGATTTGATCGACTTCGGCTCG 28633
28640 GACATGCCGAGTAAACGACGCGCAAGGTCAACGGGAGGTGTTACGGGGACAGATAC 28711
28700 GTGCTCTGTGATAGGCTCTGAGTTGACGGTGGTCTGATGTTCTCGGGGGTACGTCA 28711
28760 CCGGGTCCGTGAACCGGATTGGGGTGTGCGGATGCGCGGATGGCGGATGGCGGATTCGCG 28816
28820 ACCGCTAACAGATTGTTCGCGCGCTCCGGGAAGTTCGCTGAAGCGCTGATGGGTGACG 28821
28880 ACATGGTGTGCTGCTACTGGCTATCCACCTGCTGGGGCATCGTATATTCGA 28881
...
28948 GTGGCTACCGGGGGGAAAAATTCGGGCCAGGAAGTCTGCTCCGAACGATGACGTCGGGT

```
23989 23997 24026
24047 ACACGAAGCGTCCGACGGCCCGATCAGCGATGTTCCGGTCAGTTATATGACGGCGCAAC
24054 24063 24073
24107 ATATTGGGGCTATTGCGAGCAAAAGGCAAAAGGACTCGACTACTCGCGTTGATGGTCG
24163
24167 TCAGCTCCACGACGCGCGCCAGTGCATATCCGGGGCGCAACTACGTGATCAACGCC
24168 24220
24227 ATCTCCGACGCGACGATACCTATCGCAGCTGGTTCCAATACAGGCAA
...
24437 ACCATTTACATTTCTTTGCAAGTGTGATCATGTGCATGTGGACGGCGATGATCGTCCGGTG
24487
24497 TCAGCTCATGGAGTTCCACCTGATGTACGACGCTGGTGGGGGCCATGCCCTCTCG
24528
24557 AGCTACCGCCGCGAGCGAGCTACGAGACTTCTGCCGCCGACACACACAGTTTCAGCTCCA
24580
24617 CCCTCAGGTTGGAGTCCGCCCCAGGTTTCGGCCCTGGAGAAAGTTTCGCCGAAGGTACTAAG
24630
24677 GTAGCTTTC
...
24752 TGCTCGATGAAGACGACGCGCTCAATTTCAGTCCGTTCTGCACGGCTGCCGGCGCTCGGT
24802
24812 TCATCGGTGGCTACTAGCTCTGCGGCCCTGGCTGAACACGAGTTGACCGGTACGACAA
24872 CCTATTACGACTTAACGCGCGGACACGCGCGCACTCCAGCGGATGCCATGACCCAAAG
24885
24932 GTTGGTTACCGGCCTAATTCCGATCACCGTCCCGCTCCGCGCTCGCGTTCCGGCATG
24967 24978 24985
24992 CCGCCGAGCCGCGACACTCTGCTCAGTTCGGGCGTGAAGCTCGCCGAAGTACCCTACG
25057 25068 25105
25112 ACTTCCTCAGCAGCGCGGCTCCGCTTTCGGTACTGCTCACCGGGAGTTAACCGGTA
25172 CGAATAGAGGTGTACAGCGACGGTCCGCTACTCTTATCAACTGTCCATCTAGTCTATCC
25189 25196
25232 GCCTCAGCAGGGGAGGCGAGTGGCGTCTGTTCCCGACAAACCCGATCGCCCGGAAT
25279
25292 CGGTTGCCCGCTACCTGGCAACGCTGAAGTCTGTGTTCCAAACGAGTCCGCGAGAGCGGGC
25311
25352 AGCAGCAGAATGTTGCCTGATTCTCCGGTGGTGAACCCATCTTCGCGCGGCTAGGTG
```

```
25412 AACTCGTCCGCGCGGCGCTTGGGTTTGGTGGCTGCTTGGTCCGCTCGCCCTCGTAC
25417
25472 TGCCGATGGCGGTGCCCTTTCACCTGGCGAGATGCTCAGCGACATCCCGTCCGCGTCTGTC
25508
25532 CTGCCGACGCGCCCTCCAGCGTCCGCTTTCGCCAGATGGCCGAGCGTTCCACGAATCCG
25549 25552
25592 GCTCCGAGATATCTTGGTAGTCTGCTCACCACGAGAAAGCTTGGGAGCGCGGACG
25652 AAAACGCTTACCACACATTTGGTGGATCGTCTGGGAAACGACGCTAAAGACGTCGTGATGC
25654 25676 25687 25699 25690
25712 TGCAGGACTTCTTACTACTCCGCCATTTGCGTGAGGTGCTCGGTAGTAAAGATGGCAAG
25772 CATGGATTTCGCGATCGGTCCTCGCGGCGACCTGGGTACACCACCAAGTCTACCAAGCTT
25798
25832 ACACCGCTCGAAACGATCGTGAAACGAACCTGTGGCCGGAACACAGTTTACGCGCAACG
25837
25892 TGACGAGCCCGACGCGTGGCAGACCTGACCGACGCTGGGGTCGSGGATCGGGCTT
25927
25952 CAATCAGCTGGCGATCGCGGTGATGTTGCTAGTCATCTTGTGATGCTCATCTATCGCAACC
25962 25966
26012 CGGTTACCATGCTGTTGCCCTGGTGACGATTTGGCGCATCTTGTATGACGGCGCAGCGT
26037
26072 TGGTTCCGCGGTGTCGCTCGTCCGGCGGTCAGCCGTATCCAATCAAGCGATCGTGTGC
26072 26085 26118
26132 TCAGCAGTATGATCGCTGGTGGGGAACGATTACGCCGTTTTCTTAATCAGCCGCTATC
26143
26192 ACGAGTATGCGGGCTCGGTGAGCATCCCGAGGTCGCCGTCCAGCGCGGATGATGTCG
26238
26252 TCGGGAAGGTGATCGCGCGCTCCGCGGAACGCGTCCGAATCACCTTTCCTCGGCATGAGAT
26263
26312 TCGCCAACTCGGTGTTCTCAACGTTTGGCCCGCTCTGGCGATCGGATCGCGGTG
26352
26372 CGTCTTGGCGGGTCAACCTGCTGCCGCCATCTCTGGTGTGGCTCACCGCGCGGTT
26375
26432 GGTCTGCACCGCGCGGTGAACGATGGCGACATTTCTGGCGGGCGGCGCAACCGCAATAG
26432
```



```
21321 CTCGGCAGCCGTTGAATCCCGTTGGTGCCAGAGCACTAGCGTGGCTTCGGCGTCGGAT |-----| 21371
21381 GTCGCACAAATTCGGTTCGCTCGGAGGCAAGGCTCACCGGTGTTG
...
21529 AGCCGCGCGGCTGTTCGTGGTGACACAGAGCGCGGAGCGTGCTGCGGACGATCTTG |-----| 21579
21589 CCAACTTGAACAGCGGATTGCGTGACTGATCGGGGTGATCGATCCGAGCATCCGC
21649 ACCTGGGTGCCACCGCAATCGACGTGCACACGACGAGACCGCTCGCTGCCCTGGTGGCCA |-----| 21690
21709 GCCAACTACAGAGCGGTTCGACAGGAGGACGAAACCGC
...
21769 CGGCCCGCTGCTCCCGTCCGTTACGCCGCGCGAAGCGCGGCAACCGCTCGTCAAT |-----| 21819
21829 ACAGACGCGAGGTATGCGCTTCAGATCCCGACTCCCGGGGACCTCGAGTCTGTTGGAGT |-----| 21867
21889 TCGTCACATTGACCGGGTCGCGCGGACCGGGGAGATCGAGGTTCGGGTGACCGCAT |-----| 21878
21949 CGAGTGTCAACTTCCCGGACGTTCTGGTTCGCTTCGGGGGTATCCCACTTCGAGGGCT |-----| 21966
22009 ACCGACAGCAATTGGGCATCGA |-----| 21975
...
22107 GCCAATGCTGCTGGAGCACATTCGTACAGATCGATGCCGCTGGCGGTGACGCTCCCG |-----| 22157
22167 CCGGAGCTCCGGTGGCCGCCCGCGGTACCGACCGCTCCGGCAAGGCTTGGTAC |-----| 22257
22227 GGCCTGCACGATCTGGCTCGCATCTGCTCGGACGACAAAGTCTGATTCACTCGGGGACC |-----| 22257
22287 GGTGCTGCGGCAGCGCGGATCCGATCGACGGCGCGGATCGAGATCTTCGCC |-----| 22304
22347 ACCGGGGCAGTGCCACGCGGACAACTGCTGCACGACATGGGTGTCGAGCATGCTAC |-----| 22366
22407 GACTACGGAGCAGGATTCGCCGAGCAGATCCGAGCGGACACCGATGGGTGTCGAGCATGCTAC |-----| 22443
22467 GACGTCTACTCAACTCGCTCGCGCGCCGACAACTGATGCGGATCGAATTCGTGGCC |-----| 22491
22527 TTTGGCGGCGATTCTGTGGAGATCGGCAACGTGACATCTACGGCGACACTCGGCTCGGG |-----| 22569
22587 TTGTTCCGTTCCGCGCAACTGTGTGATGCGCTCGACTTGGCGCTGCTGACACAC |-----| 22610
22631
```

19059 GACCAGAACGCAACCGCGGACGTCGAGTTAGCATGCGGGACCTGGGCTAAACGCTGTCC
19109 |-----|
19119 TGCCGCCGTGCACACCGACCGGTCGCGAGGTTGTGACGGGGAGCAGTTTGTGACGAGCTG
19179 AGCGCGCGCTGCGCGGACGTCGCTGGCGATCAGATTCCCTATCAGCCCCGACGTGGGCGAC
19184 |-----| |-----| |-----|
19194 19203
19197
19239 |-----|
GACGACCGCGGCGGCTGGGTGTTCTCCGGGCAAGGCTCTCAGTGGCCCGGATGGGC
19239
19299 ACTGAACGTGCTGAGCGAAACCGGTGTTGCGCCGCCACCGTCGCGCGGATGGAGCCGGTG
19343 |-----|
19359 |-----| |-----| |-----|
ATCCCTAGGAGTCAGGGTTTTCGGTGACCGAAGCGATGTCGGCCACACAGACGGTCAGC
19391 19400
19419 GGTATTGACCGGGTGACAGCCACCATTCTCGCGGTGCAGGTCGCGCCTGGCGCGGCCCTG
19458 |-----|
19479 AAGTCGTATGGGGTACGTCCTGGTGCCATCATCGGGCACTCGCTCGCGGAGGTCGCGGCA
19539 |-----| |-----| |-----|
GCCGTGTCGCGGAGCACGTCGCTGCTCAGCAGCGGATTGCGAGTCATCTGCGGCGCTCG
19545 19559 19591
19599 CGGCTGATGTCGCGCATCGCGGTFAGTGGCGGATGTCATCGGTGGAACGTCGCCGGCCAA
19614 |-----|
19659 CAAGTGTGTCAGAACTGGGATTCGTGGGATCTCCGACGTCGTGCTCTCGGTGGTTGCC
19695 |-----|
19719 TCTCCGACCTCAACCGTCGTCGCGCGCGCACGACGTCGATACGTGACCTGGTGGCGGCC
19734 19743 |-----| |-----|
19779 TGGAGCAGCAGGATGTGCTGGCGCGGAGGTAGCTGTGGACGTCGCTTCACATACACCG
19818 19821 |-----|
19839 CAGGTCGATCCCATCCTGGACGAGTTGCTCGAGGTCTCTGGCCGAGGTCGATCCGACGGCG
19895 |---|
19899 |-----| |-----| |-----|
CCGGAATTCGGTATTACTCCGAACGTTGTGGGATCCGCGGACGACCGTCGTTCAAC
19900 19922 19942 19949
19959 GGGAGTACTGGGTGGAACCTGCGGTACACGCTGCGATTTCGCGG
...
20041 TGTTGCGGAGCTGGCTCCGCATCCGCTGCTCACCCTACGCGGTCGAGCAGAAGCCGCCA
20091 |-----|
20101 GTCTCGACATGCCGATCGCATCGGCTGGCCGATGCGGCGGGGGAACAGCTGCCGTTG
20120 |-----|

20161 GGTTCGCGGCTTCGTCGCGACGTCACAAACGCGCGCCAAAGGTGGACTTCTCTCTGCC
20175 |-----| |-----|
20190 20196
20221 AGTACCCTGATGGCGCTTGGTGGATGCGCATTTGCCGAGCTGGACGACCGCACCCCTGA
20233 |-----|
20281 TGCTCAGC
...
20327 GCGGTTTCATCCGCTGCTTGGGGCCCATGTCACCTGTTGGAGGAACCGGAGCGTCACGT
20377 |-----|
20387 CTGSCAGGCGGGGTTGGCACCGGGGCGCATCCGCTGGCTCGGTGAC
...
20499 CGCACCACTCTTGGCGAGCTGTCCGAGGTGCGCGACATCAAGTTCGAGCAGACGCTGTG
20549 |-----|
20559 CTGACGAGCAGAGGTGGTCTCATCGGCGCGACGATCGCGGCGCTGGGATCCTACAG
20591 |-----|
20619 TTCGACGTCGAGAGTCATCAGGAAGGCGAGCC
20595 |-----|
...
20668 CGATGTCGACGCAATTGGAGGATGCCGAGCGCGCGGTACGACACGACGACGCTCTGA
20718 |-----|
20728 CCGCGCCCCATGAGTCCAGCATGAGCGGTGAGGAACGTCGGAANAAT
...
20794 AGTATGTCGCGCTTTTTCAGGCTAGTTGCGGTGACACGCGCGCGGACGCTACCA
20844 |-----|
20854 CAGTGTCTGCGCGAGTGCCTGCTGGAGCCATCCGATCTCAGCA
...
20925 CTGCTTGATCGGTTTTCAGTCGGTGTGTTTCATCCCGAGGTCCAGAGGCGACTGTC
20975 |-----|
20985 GGTGCTGTGATGTCGCCGTGGCGGTAGGCTGGCAACTATCACTCGACGCGCAGC
21045 GCGCACTACTGCTCGCGCGGTACGTCATCGTCGCGAGCGCGGAATCGAAGCGCAT
21074 |-----|
21105 CTCGACGTGTTCCGACCGAGCCGGNA
...
21158 GCTGGCGCAGGATTTCCGAACATGAACCGCAACCGGTTGTTCCAGCAGCGATTGTT
21208 |-----|
21218 GACCATGAGTGGGAGCGGTTGAGTGTGCTGAGGTGCCGAGATC
...

16642 16679
16698 |-----| |-----| |-----| |-----| |-----|
GGGCGCCACAGTACGTAGCAGATGGTGGCAGCTCATCGAGCACCGAGCGTCCACGGGCGCTA 16757
16699 16742
16758 |-----| |-----| |-----| |-----| |-----|
CGGTCCGGCCCGCGCTCGATACCGCCCTCCCGTTGGGGTGCGAGAGATCCACGACACGCTT 16811
16760 16762
16818 GCTGGTCGCGGTTGGCAGGACTTCCCGCGGACTGTCGAATCGGTTTCGGCGTCGATGA 16867
16878 |-----| |-----| |-----| |-----| |-----|
CGGCAAGCGACTGATACGTACGAGACGTTTCATCGCGCAGGCAAGGCCACGCTCGGTTCCAC 16900 16907
16938 ACTTCGATGCCGACCGGTCCGACGACGGAACCTGTTATCCACGGCGGCCGCCACTGCG 16958
16998 CCTGCACCGCGCTGTGGGAACCAACACCAAACTCCGGCGCCCGACCGCGCGCTGCTT 17034
17058 |-----| |-----| |-----| |-----| |-----|
CCAGGCGGATCATCGTCTTGGCCGACCGCGCGCAGCACCATAGGACAGGTTGTCGC 17069 17113
17118 |-----| |-----| |-----| |-----| |-----|
CGGACGAGAGTGGCGCTCGAACCGGTCCACGCTTCTGTTGATATACCGCGCAGTTGC 17118 17127 17136 17147 17125
17178 CGGCAAGGCCCGCGATTCCAGGTC
...
17348 GCTGCTGGCTGTGAGCAGTGTGCTACTTAACTTTCCAGCGCTACGACGCTGTT 17398
17408 AGCGCATGSCAACGGCCTGTGAGTTGGCTGGATAATGTGTCTTC
...
17617 CACGGCTGGTACAGGGGACTGCGTAGTGCACACCGCAACCGCCACATCGGTCGCCGTTA 17667
17677 TCGGCATGGCTTGCCGCTCCCGGGCGCATCGATTCCCCACACAGCGCTCTGGGAAGCGC 17719 17732
17737 |-----| |-----| |-----| |-----| |-----|
TGTTACGCGCGGACGATTTGGTGGGTGAGATTCCCGCTGACCGGTGGGACGCGAAGCTGT 17737 17745 17748
17797 ACTACGACCCGAACTGGTGTCCCTGGTGCATCGGTATCGCGTGGGCGCCCTTTCTGG 17844 17856
17857 |-----| |-----| |-----| |-----| |-----|
ACGACGTCGGCGGTTTGACTGCGATTTCTTCGGCCTGACCGAGCGGAGGCGACCGCGA 17861 17906
17859

17917 TCGACCCACACGACCGGCTTGTCTGTGAAGTGTCTGTGGAGGCTATCGACACGCGGGGTG
17977 |-----| |-----| |-----| |-----| |-----|
TGGACCCGCGCAGCGTCTGCTGAATCACAACACAGGTGTCTTCTAGGACTGACACACGCGG 17984 17987 18033
18037 |-----| |-----| |-----| |-----| |-----|
ACTACGAGCTGCTGTCCGCGGATTGCGCGCGCGGGAAGGACCGTACGGAATTCACCGCA 18038 18063
18097 CCAGTACAGTTTCGCGTCCGG
...
18181 CCGGGTTACGGCTGTGCATCAAGCCTGCCGAGCCTGGATGACGGTGAAGCGATCTCG 18231
18241 CTCTTGGCGGTGGTGTGTTCTACGCTAGAACCGCGGAAGTCCGCTCTCGGGTTCCCTGC
18301 |-----| |-----| |-----| |-----| |-----|
AAGCATGTTGTCGCTTACCGGCGGTTGCCATGCCCTTCGACGAAGCAGCTGATGGCTTCG 18311 18322
18361 TGTCCGGTGAGGGGTGC
...
18384 CTGCTGCTGAAGCGGCTACCGGATGCGGTGCGCGCAGCGTGATCGTGTGCGCATCGTT 18434 18438
18444 CGTGGCACCGCAGCAACACGAGGATGGCCGCAACCGTGAATATCGCGCGCGCTCGGCGCAG 18488
18504 GCTCAGATCGGGTGTATCAGCAAGCGTTGGCTGCAGCGCGCTCGAAGCGTCAACGCTG 18527 18543 18551
18564 GGGATGTCGAAGCCACGCGCACCGGCACCGCCCGCTTGGAGATC
...
18612 GAATACCGGACGCTGGCCGCGGTGTACGGAACCGAGGGTCCGTGCGCGCTGACGTCGGTG 18662
18672 AAAACAAACTTCGGTCACTGCAGTCGGCATCGGGGCCCTGGGGTTGATGAACAAATC
18732 |-----| |-----| |-----| |-----| |-----|
CTGGCGTTCGGCATGGGGTTGTGCCGCGAGAACCTGCACCTCTCTGCCGGCTGCCTGATC 18734
...
18876 GTTTCCTCGTATGGAATGTCGGGTACCAACGTGCATGCCATCTTGGAGCAAGCGCGGTA 18926
18936 TCAGAACCGCGCTTCGGGACCTGAGCTCACTCCCGAAGCCCGGTGGGCTGGGCTGTTT 18986
18996 CCGGTGTGGCTACCTCGGCTGAGCAACTACAGTCAAGGTCACGCGCGCGCC
...

14323 GCCACCGACGAGGGGTTGCGGGGGGGTCCGACGGCTCACCTGTT
...
14393 CCGGGTCACGGCAGGTCGTCGACAGTGGTTGTCAACGAGACGGGTTATCGCCGCGG
1443
14453 AACAGATCGAGGATGGCTTGCCCTTGTGGGAGTCTGCTGGCAGTT
...
14556 TGTTCGTGCGAACACAGCGGGCACGGCTGCCAGCAACCGTGCCTGGGATAGCC
14606
14616 CGACGACGGCAGCAACCGCAACAGCCTCCGGCAACGGCCGGGGCGGTGAGGA
14617
14676 TCCTGGCTCGATGAGTAGCGGGTTCCGCAAGCGCGCCAGTAGCTCGTCAGGGCCG
14711
14736 CGGTGAACCGCAGGGACTGCTCGGTGGGCAC
...
14816 AGTCCGCGGATCCTCGCTGTGCGCGCCCGCGGCGTACAGCGATCCGCGACGACGGG
14866
14876 GCCAGTTGCTCGAGCGGTTGGCGCGTGGTCCAGCGCGGGCTTTTCGGCCCGCCGACGG
...
14936 GTTGCCACCAAGGCCAACACCGCGCGAGGATGGCGCCGCGCTGCCGACACCGCCAGC
14962 14968
14996 ATCCGAGATTGGCGAGGAACTGCGGTGGCGTCTGGCGCCGACCAACGGAACGGCG
15034
15056 GCCACGGCACCCCTTGCCAGGGGACCGGACTGAACCTGCCGGCACCGGTGGGGTCAAT
...
15116 GCCAGGGGGGATCGGGGTGCGCGACCCCGAGCGGTGGCGGAGCGCAACGGGATG
15139
15176 GTTGCAGCTGACTCCTTCGTAGGGCTCGCCGATTCCGCCACCAGGATCTCGGCTGGCG
15181
15236 CGTTCGGCGACGCGGTGCAGCGCTCGCGCGTGTGATGGCGTGGGTATCGGTGACCGGAGA
15241 15254
15296 CCGTACGGCGACAGCGACGGATCGCAGTGCCTCACACCGGATTTCGATGAGCCCTCGCGG
15302 15308
15356 GTTGGCCG
...
15377 TGTTCGGCACCAAGCGCAGGTAGTCGGCATCACCCGCGGGGTGTTCTGTGGGATACAG
15427
...
15478
...
15512
15550
15576
15706
15716
15754
15776
15911
15921
15925
15959
15981
15985
16018
16041
16127
16177
16187
16278
16328
16338
16339
16361
16398
16400
16458
16490
16518
16534
16578
16637
16638
16639

15437 CAGACCGACAGTTGTCCGCCACCTTGTGACCGTCCGAGGGTCTGTTCCGATGGCGCG
15478
15497 CCGCGCATCTGATGATCGCTGCCTGGTGGTTCGCTCGTCAAGTTCGATATTGACGTTG
15512
15557 ACGCCCGCGCAGTCCACACCTTCACTAGTAGGAAACGGGTGCCGACCA
...
15656 GTACGTGGGTGAAGCCGCGGTGCCGCGCATGACCCGCACTAGGGTGGCGTCAAGC
15706
15716 GGTCCGGGTGACGAGCGCGCGCTCGGCGCAAGAGCGGATAGGTCATCTTCGATC
15754
15776 GCGCGCGCGCAGCGGAAGGTTTGACCTGTTACCA
...
15861 CGGCAACAGCTGGGCGCAACCGACTGCTCGCTGACGGTGGCGCTTTCAGCATGTG
15911
15921 GAAGCGCGCGGTGATCGATTCGAAATCGCAGACCAACGCGCCCAACCGCGCCCCA
15925
15981 AGACGGCTCTCGGTGTCGAGGATGTCGCGGTGCGCGGATCTTGGATTTCGACACG
15985
16041 CGCACAGTCTGTCTACTGGCGAGGTCCGACG
...
16127 GCGGTACGCGGTGCGCTCGCGCGGATCCGCGCTTTGTGAGTGCAGCGCTGAAG
16177
16187 TCGGTCAATACGTTAAACCCAGTCTTGGCATCGGGCGCAATTCGCT
...
16278 CGTAGCGAAGCGCAGGCGCTGCAGCGAGGTTCGGTTCGCGCATCGCTCGACGCTCGAC
16328
16338 AGCGGATCTGATTCGCTCGCGCTCGCTCGATCCAGATCCGGGTGTGACGCCACGCG
16339
16398 CCAACGACATGCTGCCACCTTTTGGTCGATGAGGCCAGCATGAGTTCGGCGAAGCGG
16400
16458 CCGGTGGGTGCCGATCCAGGCTGCTTTCGGCGCTCGGTGGGTGAGATAGACCAACT
16490
16518 CTTGGTAGGAGCCAGGTCCTTCCCTAACCAAGCGGTGTTCGGGATCAGCAAGTCG
16534
16578 CGGTCCCGAACAGCTCATGACGAGGTGTGCTGCCACGGGTGAGCTCTGTGGCGGGG
16637
16638 TCGCGGTTAGCCGATCAGCGGCTCTGCGCTCCGAGGACCGACGCAACGCACTGACCA
16639

12157
12213 GCCGCGCAGCAGAGGTATTGATGCAAGTGGCGCGCATCGAGCGCGAGGATCTGCGAA
12243 12253
12273 GGTGCGCTTCGTGATACCGCGATAGGTGTCCAGTGACTTGACACGCTCGGCGCGCAGCGC
12333 GGTGACTTGGGTATACAGCGCGGCAACGTCCTCCGTAGCGCGGCACACGAGTTGACGCGC
12372 12390
12393 GATATCGGCGCTGTGCTGTGGTCAGCGCACTGCGCGCGCGCGGCGCTGGTGGGCTGAG
12395
12453 CGGTCGGCGATCCAGCGAGCAGCTCGGCGCGCGCGCTGTGCGAGATCGGCTATACCTT
12459
12513 CTGTGCGGCGCGCATCTGAGCAGCGTGTGCGCGCGCGCGCTGTGCGAGATCGGCTATACCTT
12541
12573 CACCGTAACGTCGCGATAGATCTGCTGATGGTTGCGGTCCAGAAATAGCCGATGGAGCG
12580 12629
12633 ATCGAATGGCGGCGCCAACTCGTAGCCCGAGCTGGCCTGTTACACAGCGACACCAAGAT
12634 12679
12693 CGCGTGGGTAGCGCCATCCGGGAGCATAGACAGCGCAAGCCGGATGCTATGCAACTAGG
12703
12753 TGCATATTGACCGTGTACGCGCGCATGTGCCAAGTGGTCGACGTATGTGCAAGTC
12807
12813 TAGTATCAGTAACGACGATTCCTCAGCAGGGCGCGGAGGAGCCTT
...
12866 GAGGTGACACAGCCCCACACATAGCGGTATCTGGAAGACATGTTGAGGAGACGTCGCTG
12916
12926 ACGTACACGATCGCGCAACCTGTGTGACATCAAGCACAAGGCATGCATTGAGGAGTGC
12935
12986 CCGTGTGATTCATCTACGAGGCGCGCGGATGCTGTATATCCACCCGACGAATGCGTC
13007
13046 GACTGTGGGCTTCGAGCGGCTGCCCCGTGGAAGCTATCTTCTACGAAGACGATGTG
13097
13106 CCGGAACAGTGGAGCCATTACACCCAGATCAAGCCGATTTCTTCGCCGAGCTGGGATCG
13136
13166 CCGGGCGGTGCGGCCAAGTTGGCATACCGAGAAGACCCGCAAGCGGTCAAGGATCTG
13167 13199
13226 GCSCCGCAGACGAGGACGCGCTAGCGGGTGGGGGAGCACCCGCTCGCGCGGAGTGT
13230 13241
13255

13286 CGGCGTCTCTGCCCCGTCTTCCCTGGGACACCTTGGCCGACGCGAAAGCGGTGGCCGGG
13327 13332
13346 CCCATCCGATGGCATCGTCGACCTCTCCGTGGCACTCCGGTCGACCCGGTCGACCGC
13360
13406 TGATCCAGAGGCGCTGGCGGCGAGTGGCCGCCCTGGGTATTCGCGCGACCCGCGCA
13416 13452
13466 CCGCACGTTACGTGAGTCTGTGGTGGCAGCGCTGGCTGCGCGCTACGGCATCACAGGC
13494
13526 TGACCGAGGCGCGTGTGGCGGTTATFCGCAACCAAGAACTCATCCCTGGTTGCCGA
13570 13584
13586 CGTTGTGGGCTGGCGGTCGGATCTGTGTCGTGCGCGGAAATTTGGCATATCCGACTT
13589 13615
13646 ATGACGTGGCGCGCGCTGGCGGAAAGCGGGTGTCTGCTGCGGATGCGCTGACCCAGC
13648 13654
13706 TGGG
...
13766 TGCTGGTGTGCAACATTTGCGCAAGTGTGAGTGGCCCGGCGAGAGCGCTTCTCG
13816
13826 TGTTTCCGACAGTGTCTACCTGGGATTGGGCTGGGACGCGCAACCGGTTTCGGTGTAC
13861
13886 ATCCCTCGTGTGGCAGCGGACACACCGGTTGCTGGCTGTGCACTCACTATCGAAGA
13903
13946 GCTCATCGCTCGCGGCTACCGAGCGGTTTCTGCTGCTGGTGACCTCGAGATCGTTGCCG
13950 13978 13996
14006 AGCTACTAGCGGTGCGCAACACGCGCGGATGATGTGCGCGCGCGGTACAGCGGCTA
14046
14066 TGTGGCGCGCTGGACGACGACGCGCACGAAAGGCAACAGCGGAGCGCTACGCAAC
14080 14083
14126 GCGGTGCGCGCTGTTCGCCGCGCTGGGCTCCCGGGTTTTCGGGTGCACTATTCGAGC
14145 14182
14186 CCGGATTTGATCTATGGGCCACTCGCGCGAGCGGTGCCCGGACAGTGGCGC
14187
...
14263 GTGGCACGGGTGATTTCTACGGCCCGGTGGGGCTCAGCAGCTGCGGGTGGCGGTGACG
14313

9964 9969

9987 CGTCTGCGGTGAGGGTACCAGGCGCGGTGGCTTGCTTCTCTCAAACTGGTTATCGGCGC |----| |--
10008 10044
10047 GACACTGCGGCGCCATACCAATCTGCAGGTACAGCAGCGATGAACAACAGTTGTTTACAGC |----| |----|
10049 10081 10092
10107 CCGAGAAATGAGTTTATAGCCTGCGCGCAAGTTCGGTGCCCTTGCTTGATGCGCGCGCTGG |--
10165
10167 CGTCCAACTCAGCGCAACCGCCCGCCCGCATGATGCGGGTTAATGCCGT ---| 10170
...
10289 GCCAGAGCTGGGGCCCGCTTGGCTTCGGGCCGAAGCTGATGTAGGCCGTGCTCCTTG |----| |----|
10339 10342
10349 ATCTGTTCTGAGTTACCCCGCAGACAGCTGATGAACGCCCTTGGCCTTCAACGAGCGCCCG |----| |----|
10381 10397 10400
10409 TGGACCCATCCGGTGGTCTTCCCGAGCGCTTGCCCTGCGGGCTTTGGTGCCTGCAGT |----| |----|
10469 AGGTACACCTCACGGGCGGCGCGCGCAGTGGAGTCTGTCACACGCTCCGCGGGCTCT |----| |----|
10490 10505 10511
10529 CGCGGATCAGCGACCCCGCCCATTCGGCCTTCCACTCTTTGAGGTTGAGGTTGGGTGAGGAG |----| |----|
10537 10609 10630 10636
10589 TCGGTGACCAGCAGTTCGGTGTGCTCGAAGCCCAATGCCCGCGGCGCGCAGCAGCCACG |----| |----|
10649 GTTCGCCCGACCGCTCTGACACCGGTGATGGCTTCGGCGTAGG |----| |----|
...
10706 TGGTCGATCGCGGGATGCGCGAATCGCGGTGCCACGCGGTGGCCAAAGCAGACCTCG |----| |----|
10756
10766 TCGTAGCCGCTCAACTCTCTGGGCGGCCACCCGAGTGCCCAAGTCGCACCTCGACACCGTGT |----| |----|
10826 TTGGCCAGAATCGTCGAGAAATACCGGATGGTTTCGCTGAATTCCTCTTTGCGCGGAATG |----| |----|
10835
10886 CGGCGGCCATGTCAAACCTCTCCACCGATAAAGTCGTTGGCCCTCGAACACCGTGAACCCG |----| |----|
10918
10946 TGACCCCGTTGCGGGGCTTGGCGCGCGTGGCCAGCCCGCTGGTCCAGCCCCGAGCAGCG |----| |----|
10960 10999
11006 GCCACCGAGCGGCGCGGCTCGGGGACAGCACAACCTTCTGCTCTCGC

...
11138 GTGTTGATTTGCTCGACCCGATTGGACTGCGCCTTGAGCACCAGTCCGGGTCCTCAGC |----|
11188
11198 ATCGGCGGCGCAATTGATATCAGCGGCACCTGGGTTTCGGCCAGAATCCGTTCCGCGGCC |----| |----|
11258 TCGGCGCATGTTGATPCGGTTGGACGCCCAACCCACCGGGATAGTGAGCTTTCGGCGACGCGG |----| |----|
11279 11308 11314
-|
11318 CTGCTGATGTCGACAAACGCGCGCCCGCACTGAGGTGACGATAGTGGCAACCCGGGCC |----| |----|
11319 11356
11378 TCGTGCCAGCCGGAAGCCGGAGTTGATGATGGTTGCGCCTTCCACTTCGCTTGGTTGCC |----| |----|
11438 AGCGCAGCAGATTTTCATCCCAACTCTGGCCTTCTGCAACGTAGTCGGCCATTGACAGCCG |----| |----|
11443
11498 T
...
11512 GTCGCATCCGACGCGCGCGGCTGCGTCGGATGATCTCGACCGGGAACCGGACCGGTT |----| |----|
11562
11572 GGCGGCTGTCCGCGCCCAACGAGTCGGTGCCTTGTGGTGGCGCGCGCCAGGAAGTGTAT |----| |----|
11615
11632 GAGCAGATACCTTCGCTGCCCATGATTTGACGCGCTCTGAGCGGCATCGCGGGGCCAA |----| |----|
11662
11692 CTGCGCAGCGGCGGAATCCGCGATGGTCTCGCTTCGACCCCGGAGCCGATAGTGTCTG |----| |----|
11720
11752 CGGACGAACCGGGTGATCGCGCCTTGATCGGCGAGCGCTGACCCGAAGTGGGTGTA |----| |----|
11771 11788
11812 GCGTAGCGTCCGGGTGACAGATTTGCACGAGATCTTTGCACCCGAATCGTGGACCGC |----| |----|
11817
11872 C
...
11933 AGCAGCATCCGGTGGCGCGTACCCACCGGTGATGATCAGCCCGGAGCCCGCGC |----| |----|
11983
11993 CGTGCAGTTCCGGGAAGTAGTCGGCGAGCCGATCGATATGCGCGG
...
12093 TGTAAACCAAGTCCACGGGACAGCAGATTTGGGTATGGATTTGTGATCGCTTCTC |----| |----|
12143
12153 CTGGAGCGCTTCAGCTACTTCGTGAGCCAATCGATGGCACTTCTTTCGGCTCGGATTCC |----| |----|

```
7803 GCGGTGGCGCGGCTGGTCTGGGCCCCGGGCATCGCATCCAGGCGCCGCCGAACCTGGTG      7793 7798
      |-----|
      7844 7849

7863 TCTGGCAGCAATGCCGGCGCTGTTGGCGCCGGGTGACGACTGGGGCGGTCTCTCA      7880 7885 7890 7895 7902 7907
      |-----| |-----| |-----| |-----|
      7866 7871 7880 7885 7890 7895 7902 7907

7923 CCGTTGACGCCCGACCATGTCAACCCCGAACCGCCCTGGCCCGCTTTGGACGAGCTGGCG      7928 7933
      |-----|
      7928 7933

7983 G
      |-----|
      7928 7933

...

8043 GTACAGGCGGCGCGCTGGATCGACCCCGGGTGGCGGACATGTGGTGGCGTGGCG      8093 8098
      |-----|
      8093 8098

8103 GATCGGCGACCGGCTGGCCCGGACGTCAAACCGGTGGCATGCCGTGGCAGGAGCCC      8108 8113
      |-----| |-----|
      8108 8113 8127 8132

8163 GACGACGTGGCGTCTTGGGGCGGTCGATCTGGCGCAGCGATCGACACTCAGGGCGCG      8163 8168 8176
      |-----| |-----| |-----|
      8163 8168 8176 8201 8206

8223 AATACCGAGTGCAGCAGCTGGCCAGCCCTTCGGTGACTGGGAATCGATCGCGAG      8238 8243 8250 8255
      |-----| |-----|
      8238 8243 8250 8255

8283 CAGGTGCACGAGCTGGCGGTCCG
      |-----|
      8238 8243 8250 8255

...

8343 CTGCGATCGGCGGAGCGTGCCGCCCGCGGTGCACCGACGGCGAGTATCTGGCGCTGCC      8393 8398
      |-----|
      8393 8398

8403 ACCGCCGACGCTCCTGCGCTGGAAGCCCTTGCCGCACTGGCTGATTCGTTGGCGCGCAT      8469 8474
      |-----|
      8463 8468 8469 8474

8463 GTCTCGCGCAGCAGGTGACCTTTGTGTCAACCGTAACATCAACTCACCAACATCTGC      8463 8468
      |-----|
      8463 8468 8469 8474

8523 TACACCGGTTCCCGGTTCTGGCGCTTCGCCCAGGAAAGGTGACCGCAGCGCCTACTCG      8565 8570 8571 8576
      |-----| |-----|
      8565 8570 8571 8576

8583 CTGTGCGTCGAGAGGTGCGCGCACCGGCATGGAGGCCACGTCGCGCGGGGCCACCGAA      8598 8603
      |-----| |-----|
      8598 8603 8625 8630

8643 GTATGATGCAGGGCGGTATCGATCCCGAGCTACCGGT
      |-----|
      8598 8603 8625 8630

...

8710 TCAAGCGCGGTTGCCCTCCATGTCATGTCACGCGTTTCCCGGATGGAGATCGCCACG      8760 8765
      |-----|
      8760 8765

8770 CGGTCAACCAAGCGGGGTGAGCATTCGCGAGTGGCTGATCGGCTGCGCGAGCGCGGGC      8774
      |-----|
      8774

8830 TGGATACCATCCCGGTTACCGCCCGGGAATCCTGGACACAGAGTTCGCTGGGTGCTGA      8865 8870
      |-----|
      8865 8870

8890 CCAAGGCAAGCTGCCGACGTCATTTGATCGAATCGTGACGACCCGCCACAGAGGTGG      8906 8911
      |-----|
      8906 8911

8950 GTCTCGGTTCATCATCGACGATGATGACGGCATGTGACAGTCCACGGCACTGGGTGG      8966 8971
      |-----|
      8966 8971

9010 CCCATCTTAACGTGCTGCGCGATATTTCAGGACCGTACCGCGGCTTCACCGA      9011
      |-----|
      9011

...

9205 CCAGCTGGTGAAACTTGGAGTGGCGGCACCCAGGTGATGCTCAAGGTGGCGCAAG      9255 9260 9261
      |-----|
      9255 9260 9261

9265 ACCTGGCGGCACGCTGATGGAGGAGACCATCTCGCGGATGCGCGTTCCGAACAGGAT      9266
      |-----|
      9266

9325 CGGCCAAGACGTCGCTGAGTGGTCGCGATCGCGCAAGGCATCGGCCCGCCGCGCGCC      9336 9341
      |-----| |-----|
      9336 9341 9354 9359

9385 AGCGCACTACCACATACGCCCTGCTTGGCGCTAGCCCGCGCAGCATGCCGGTTCGCGG      9424 9429 9427 9432
      |-----|
      9424 9429 9427 9432

9445 GATGGGCCCCGTTGAGAGAGCGGGCAATCTGGCCTAGCCCCCGCGCAGCATGCCGGTGC      9486 9491 9489 9494
      |-----|
      9486 9491 9489 9494

9505 GGGATGGGCCCCGTTGAGAGCGGGCAATCTGGCCTAGCCCCCGCGCAGCATGCCGGTGC      9548 9553 9551 9556
      |-----|
      9548 9553 9551 9556

9565 GCGGATGGGCCCCGCATGGGCTTAATAGTTTTCAGGAGC
      |-----|
      9565 9566

...

9687 GGTGCGGTGGGGGAGGAAATTGGCAATAGGACTGGCCACCGGTGAGCGTTGAAC      9737 9742
      |-----|
      9737 9742

9747 TGTGGGCAGCCCCCGGATCCGTCGCTTGAGCGCAGCTACTACTGCCCGTAATTGCAG      9842 9847
      |-----|
      9842 9847

9807 GTGGTGTTAATACGCGCTCCACGGATCTCGGAGGAGCCCCCGCCCGCAGCGTCAAC      9842 9847
      |-----|
      9842 9847

9867 GACATTGCCAGCGGCTTACACCGCGCTCAATGGGTCAACGACGACCTCATTTATGGA      9869 9878 9883 9890 9895
      |-----| |-----|
      9869 9878 9883 9890 9895

9927 CACCTTCCCCAAACTATTGTCACCGTCTGTTAAGACGGCGACACATCTGCCCGCGGTTGC      9950 9955 9961 9966
      |-----| |-----|
      9950 9955 9961 9966
```

5045 TCGCCGGCGCTCCGGCGCGCCCGCGATGAACGTGGCTGGAAACTGATCTGCTCGCG
5062 5067 |-----|
5105 GTCGCGGCTTCAC
...
5414 AGTGGATTGGTCAATGCTGCCTGGTGTGGGCCCGAGCGGATGGACCAATCGTCGGCG
5464 5469 |-----|
5474 GTGCCCGGCTGCCCTGTGGCGTGGCTACTGACGGTAGCGGAGTACGCTTCGGCGGC
5515 5520 |-----|
5534 CCTGGCAGCATGTCAATCTTGGTGGCCAGGCGCTGTGTCTACGGCGGCTGGTGTTC
5537 5542 |-----|
5540 5545
5594 GT
...
5932 CCCACCCCATGCGGTGAGGTTGAAGCCGTACTGACCCACGCTTCGACGGCGCTGTAG
5982 5987 |-----|
5992 AGATTCTTGATTGGCTTGCCGATCGACTCGCGGCCAGAGCCGCGCGCGCGCG
6037 6042 |-----|
6052 GCCGCTCCGGCGCGCAGCAGGCGCCGACGCGCGGACAAACCGCTGGGCCACCGGTC
6075 6080 |-----|
6112 GGTGGAGCTGCAGTGCACCGCTGGCAGCGGCATTTGCAGCGGCGCTACCGCCAAAGTCA
6153 6158 |-----|
6172 CTGAGCGCGCAGTGCAGGAGAACATGCTTTGGACCGGCTCAGCGGCTTCGAGCCGCCA
6246 6251 |-----|
6232 CCGCTCAGGCGTGGCGCTAAATCAGTGGATGCCCGCCAGCAACGCTCTTGGCGGCGATTC
6240 6245 |-----|
6246 6251
6292 ACGGCTTACCATGGTCTGGCGGCGCTGGCTCGCGGTTGCATACGCGTCTGCATTT
6294 6299 |-----|
6315 6320
6352 GCTCTCAGCGCTGCACGAACCTGGTCTGGGAGGCTGTCTATCTCTCCCGGCTGAGTGC
6358 6363 |-----|
6412 TGATACCCCTGAGCATGCGGGAAGCAGCGCGCGGACCTGAGTCGAGACCTGCTCGCG
6444 6449 |-----|
6472 GCTGGCCAGGACTCCGGTGGTGGGAACCGCGCGCAACCATATGGCGCGGTTAAGATC
6519 6524 |-----|
6532 GAACCGATACCGGCATGTCGGCAGCGCGCGCGCAGTGCCTCTGGCGCGCGCAACACA
6577 6582 |-----|
6592 AACGACATCTGTACCTTCTCTCTGTTCACCGCGCGGCTGTGCGCGGCGGCTTGTTC
6592 6597 |-----|
6634 6639

6652 |-----|
AGCGCTGGCCTCTCACGGATGGTATCGCGATCGGCTGTGACCTGGCGCTTACTCCACCA
6653 6658
6712 AACCGTTGGTSCCGACGGTCGACGGGTGCCGAGCTCGGCTGGCGTACTGTTGCGCT
6755 6760 |-----|
6772 TATGGCCCAAGGTTGGCCACGATCTCACCTGGTGGGCGGTGCGGATGATATCAGATTG
6775 6780 |-----|
...
6870 AAGTCCACCGCTCTGCCGAGTCCCGTTCACCGCCCCAGGCAATGCCCTAGCGTTGGCG
6920 6925 |-----|
6930 CGGCTACTGCGACGGGCCGAGATGGTCTCACGCTGAACGTGGATGAGCGGCCCATAGCG
6986 |-----|
6990 -|
ATGACCGCAGCGGTGACGAGCTGGCGGACCTGTGCGCGAGCGCGCGCGGCTGCGCAT
6991 7034 |-----|
7050 GCGGCTCTGCTGTCGCGCGCCGCGACGCGGCCAGGTTGGCGATCAGTATTCG
7094 7099 |-----|
7110 CGCAAGGTGTTTATCCCGGTACCCCGTTATGCGCGGACA
...
7186 AGCTACGCGGCCAAGGTTCCAGCACGTATATGGAACCGCAGAGATCTCTCAGCTTGCC
7236 7241 |-----|
7246 GCCGAGTGCCGAATTCGGTTGAAGAAAGCCTATTACCTCTCGGTGACCGTCCGAGG
7274 7279 |-----|
7306 CGGTTGGCGCCAGGCAGCGAATGGCTCGCGAAGCGGCTATGACTCCAGTTGTCCT
7312 7317 |-----|
7366 AC
...
7419 CTGAACCCGGGTGTGATGAGCTGGTCGAGATGTGCGGGCTCAAACCGGTGGCGCGTCG
7469 7474 |-----|
7479 ATGGGCATGATGTGGAGACGACCTCGCGAGCGGTGTTTGAACCAAGGGCTCGCCAC
7496 7501 |-----|
7539 TAGGGCAGCCTGACAAAGACCCGGGCTGCGGCTGCGTGTCTGTACCGCGCGCGG
7587 7592 |-----|
7599 TTGTTCATTCCGTTTACCACCGTCTGTTGTCGGCATCGCGAGACGCTATATCGAGCGC
7643 7648 |-----|
7659 GCCGATACGTTACATGCGATTTCGCAAGTCGCACAAAGGAT
...
7743 GAACACACCGCGATGGCGCCTTCCCGATGCCGGAATCGAGGATTACCTGGCGACGTT
|-----|

...
2478 CGATGCTTCCATGGTCAAGCCCGCGTATCCGGCGGATTCGAGTTCTGGGCCACGGTCGCAT
2528 2533
2538 CGTAGAGCGCAGCGCTCAAGCACTCGCGCGCGCGCGGTACGGGT
...
2686 CCGCGCGGGTGACCGCAGTATCACTCCGGAACGGGTCTTTGCGCGCAGCGCGCCACA
2736 2741
2746 GCGCCCGCGCCAGCGCTTGCCAAATCCCACTGGGCCCCACGAGGTTCCCGCAGGGAC
2750
...
3050 GTATCGACGAGATTCGCGCGCTGTGAGCGGTAGTGCCCTTCGTCAACAGCGTCCACA
3100 3105
3110 GCTTCGAAGCAGATCTTTGTCGCGCGCGCGGGAATGACCCACCGTCTCGGACACGG
3130 3135
3170 AACCAGTCAGTTTCATAAAGTCTGGCCAGCGCGCGGTTGGCGATCGGCCCGCGGGCG
3210 3215 3222
3230 AGAACGGTGTGAGCGGTGCGCGCGCGCCGCAATGTGGCGTAGAGGTTTCATCGCCTCGC
3280 3285
3290 CATCTTGGGACCACTACTGGGCATACAGTGATCGAGGGCAAGAT
...
3394 TTGCTCTCGATCACCGCAGCGGCACGTGTCTGCGTGGCGCGTTTCATAGGCGCCTGCC
3444 3449
3454 ACAGCGTGTGATGGTTCGCGCAGCAGTTACGCCAGGGAAGGTCGTGACAATCCAGGCC
3456 3461 3471 3476
3514 CGAAATGGGCGACCGCAGTGCATGATGTCGACCATGGCCCCCGCCACGATGTGATC
3522 3527 3541 3546 3547 3552
3574 AGCCCGTTGATCTCACTCTCGAACCGACTGGCGCGGTAGCTCAGCTGTTGGACAGATT
3634 TTCCAGCGTTTCGGGCTACTAGAAACGGACGAGCGCTACCTTGGATGTTGAGGGATTG
3639 3644 3666 3671
3694 AACTCGGGGAAAAATTGGAATCCATTGTCGCTCAACCGCTGTCTAGGTGGAGGTG
3724 3729
3754 CCGCGGGTTGGCTAATTCGGTGAGC
...
3799 TCTAAGTGTTCGACAAATGACTCGTGATCAGATGGGCTGGCGCGCATCGCCTTGT
3849 3854

3859 ACAGCTCGCGTGCATGGAAACAGCATCAGCTCAGCATGGACACAAGATCGTGGCGG
3888 3893
3919 GGGATTCACATTTGGTGTATCAGCGCGGTACCCCGTCAATCATGGCGGCTCATGTCACCC
3962 3967 3974
3979 CGATCTCGTGGAGGTTGGCGCGCTTCCCAATCGAATCGGGCGG
...
4035 AAAGACCGGTGATCTCTCTCCACTGACGTGGTCTGATGTGGGGGTACGCGAGCT
4085 4090 4088 4093
4095 TGGGTTCCGCACGCGCATTTAGACGGAATCGTTCACATAAGTATTTTACCATAACGGC
4123 4128
4155 TTCGTCACAAAACGGTAGCATTCCTGTGAGGAATTTTTCGACGCTCGCCCGGTAGGG
4172 4177 4197 4202
4215 TGCTTCCATGCTGAGACGCGCGCGCTCTCTTTGTCATGCACACCCCGCAGATGAGAG
4230 4235 4264 4269
4275 CCTGAGCAACGGCGCAACCATCGCGCACTACACCTCCGTCGCGCACAGGTCCATGTCGT
4330
4335 CACGTGCACCCCTGGGTGAGGAGGCGAGGTCAATTGGCATCGCTGGGCTCAACTCACCGC
4369 4374 4373 4378
4395 CGATCATCGGACCAACTCGGTGGCTACCGCATCGCGAGCTACCGCGCGGTGCGGCGCATCCGGCAT
4455 GCTCGGGTCAAGCCACCGATCTACTTGGCGCGCGGTGCTGCGGCGCGCATCCGGCAT
4457 4493 4498 4443 4448
4515 GCGCGCACAGACCAGCGGAGTCAGCGGAGATT
...
4685 CCGCGCGGTGCGCGCAGCGGTGTTGGTCCGCTACCGCAGATCACCCCGCGCACCGCT
4735 4740
4745 GGACGGTCCGAAGTTCTACTGACGGTCTTGGGTCTGAGCGCGCTCATTTTCGGCGCGC
4805 GAGCCCTGTCCCGCAGCATCTCGACCCGAATGGGTGTTGCCCGCGCGCCAGCAGATTG
4819 4824
4865 CATTCGGGTACTCCGACGCGGTATCGACGCGTCTGCGAGCGCGATGAGCAGCGCGAG
4879 4884 4891 4896 4897 4902
4925 CCGCCAAGGTTGGGCACTGGTGCCTATGCCACCAAGTTGTGTCGTCGCGCCGACCGCGC
4966 4971
4985 GGGCCCGCGCTTGTGAAACACCTGGCACTGCCCATCTTGGCGCGATGACCATTCGTGC

MtV005 Length: 37840 February 21, 2002 16:59 Type: N Check: 8654 ..
Found using 'seq2-3' (pappu403.key)

```
1 1-----|
  AACGCCACTGCGAAATCCAGCCCGGCTTTTCGCAGCCGGGTTACGCTCGTGGGGGT
  1 6
...
241 GGATCGACTGGATCGGCCCGCCGGTTCAACCCGCATTTGCTGCCAGCTTCGCCGGTG
  291 296
301 1-----|
  ACGCCGCAATCAGTCTGTGTTGGTTGCCAGCGCGTTCGGGCCGATTCCGGGATCC
  301 306
...
428 ACCCGTTCACGGTGACCTACAGATCCGGCCCGAGCGCGAGTGGACGGACAGCCCGCA
  478 483 487
488 1-----|
  TCGCCCGCAGGACTTCTGTGTAATCTGTGTCAGCAGATGGTCACACAGCGCGGCTCGTCG
  492 496 501
548 1-----|
  ACCCCGCGGATACCACTGATCACCAGTGTCCAGTCTGTCAGTCCGAGGCGGTAAGCAGCGG
  582 587
608 1-----|
  TCGTTTACGTTGGACAGCCCTACCCCGCTTGGCGTGAGTTGTTACCGACATCCTCCCGG
  612
668 1-----|
  CGCAGATCTGTCAGGACATACACAGGGGGCTTCGCGTCCGGTTTGGCTTCGAGCGCTCCGG
  673 678
728 TGACAGGTGGACAGTTTCGGGTGGAACAATCGACCCACAGCGCGATGAGATCCTGATCG
  717 722
788 1-----|
  CCGGCAATGACCGTTTACTGGGGCCCACTTCCAAACCCGGGCATCAATCTTTCGCCCGGG
  789
848 1-----|
  CCGGGCGCGCGCGCTGGCGGATTCGGTACGTAACGGAGACACCCAGGTGCGCCAGG
  852 857
908 TGCATGGTGGCTCGGGGCGCTTCGCCCACTTGTGGCCATCCCGA
  898 903
...
959 GGACCGCCCGGATCGTGTACACCGCGGGTCAATGAGTTCACGCTGCGGGCAAAACGTTCCCA
  1009 1014
1019 AGCTGGCCGACACCCAGGTTTCGAAGGCGATTTTGGGGTTGCTGGACGTGGACCTACTTG
  1044 1049
1079 CCGCGGTGGCGCGCGCACACACCGTCACTTGGACCGGCGCAGATTCGTTCCG
  1087 1092
1139 CGAGTGACCCGGTTATGTTCCGACCGCGCTCCGCAATGAGCAGCGCGCGCGCTGG
  1183 1188
```

```
1199 GTCTGCTGGAGGATCGGGATTCAGGTCGACACACACAGCTCGGTGCGCGCGCCGCT
  1245 1250
1259 CGGTCCCGGATTCGACGACCAAGTTCAGGTCGACACCGCGCGCGGAGTTCATCCGGCGCC
  1272 1277
1319 GGATCAGCAAGACGCGGAACAGTTAACGCTGTGTATCGGGGTGGCGCGCAAGATCCCGA
  1344 1349
1379 CCTCGGTGGCGTGCACACACATGTCGCGACACAGCTGCGCGCATCGCGCGCA
  1390 1395
1439 CTGTGCTGGCTTAGACCCCGTCACGCTCTATCACGACGCGCTGAACGCAATCGGGTAG
  1446 1451
1499 ACGCCATTGTGGCTGGCGCAAGCCGCGGAAACCTGGCGAGCTGTGGCCTCTCGTT
  1503 1514 1519
1559 ACGGCTGTCCGCAATTCAGGCGACAGCTCCCGCTGCGAATGCGCGACACGCGCC
  1578 1583
1619 CATTCCCTCCCATTTGGCCTTACGCGTCCGCGCGCGACACCGGACACCGCCACCAA
  1581 1586
1679 CCGCGCGCGCGCGCCCATCGACCGCGCGCTGGTAAAGCGCGCTGAATCTCACCG
  1680 1685
1739 GCATCTGCACCGCAGCATCCAGTCGAACATCGATGCCGCACTCAATGGCACCAAGAACA
  1719 1724
1799 TCAACGACCTGATCACCGCGTCCGAGCTGGAATATGTCACCGCTTGGCGGA
  1801 1806
1859 TCCTGCAGGACACCAAGATCGTCCGCGCGCGCGAGCTGCAGAACGTCAGCTGTCTG
  1876 1881
1919 GTGCGGTGCAGTGGGCATCGTCCGCGCGCGCGCAATGGGTGAAGACCGGCAATAGC
  1936 1941
1979 CCTGTACCGCGCGGAATCGTCCGCTAGCTCTCCGCGGCTTCGCGCGTGGTGAAGATC
  1997 2002
2039 ATGGCGTTCATTAATGCGTGTGAGCTCTCACGGTCCGCGGGGATCCGGTAACAAGA
  2041 2046
...
2222 ACCGAGTTGATGAAGATGGCGCGCAGCTCGGGCTCGTGTAGGAGCTGGCTGAGCATTTCC
  2272 2277
2282 ATGCTGGGAGGGCGCTCTTGGCGGCGAGGATTTTCGACGTTGGCGG
```

| | | | | | | | |
|------|-----|---|---|---|---|-----------------------|--|
| gene | CDS | complement(31607..33223) /gene="Rv1186c" complement(31607..33223) /gene="Rv1186c" /note="Rv1186c, (MTV005.22c), len: 538. Unknown. Contains helix turn helix motif from aa 488-509(+2.81 SD)" /codon_start=1 /transl_table=11 /product="hypothetical protein Rv1186c" /protein_id="CAAL5863.1" /db_xref="GI:2695970" /db_xref="SPTREMBL:O50442" /translation="MRTAGVGLGQLLALDATTVSVSLVDAPRGDLPLVASTALIDSDDV RLGLAAAGADYFFLIGVTDDEAVRWDDQARAPVAFVFKHPSDSVAVAGAVRAGS AVAVAPPRARWERYLHVNLVHGHGDRADPTDSDGTDLFGLAQLADRLHGLMSIED AQSHVLAYSANDEADELRSLSTILGRAGPEHLQWLGQWIFDALRGREVVVRAERP ELGLRPLAIGIHQPGVGALRPVPFAGTIWVQGSQPLDADDEMLRGAALVAARIMS RLATQTHALRVQQLGLLAELNATTAPVDVSTIARELGVAAEGNATLIGFDTEANRD TAVRHVLVDVMSASAFRHDQAQAANGSRIYVLLPQTTPTRAVTSWVGTISALRA ELGVLAATAGVAGLAENVARVEVDVLESAERHPILGOVTSILAEARTVLLDEI VTLVGTGDRQLVDPRIRDLGAQDPVLAQTLRAYLDAFGDICAARSQVHPNTVYRIR RIEQLLSTIGDPPDVRLLFSLGRAMERTA" complement(33233..33236) /note="possible RBS upstream of Rv1186c" 33308..34939 /gene="roca" 33308..34939 /gene="roca" | /codon_start=1 /transl_table=11 /product="fadD21" /protein_id="CAAL5862.1" /db_xref="GI:2695969" /db_xref="SWISS-PROT:O50441" /translation="MSDSSVLSLLRERAGLPQDDAAFTYIDYEQDWAGITETLTWSEV RFRTRIVAEHVRHCTTGDRAVILAPOGLAYIAAFLGSMOAGATAVPLSPVQIGSHDE RVSRAVADASPVSILTTSAVAEAVAEHHRPNTNVGPILIEDSLDITGNSPSRVKD LPSAAVLQVTSSTPAGVWISHRNLQANFOQLMSNYFGDRNGVAPDPTIVSWLFP YHDMGLVIGLIAPILGYSRSELSPLAFQRPARWLHSLANGSPSWAANPYAFELAV RKTTDADIEGLDGLNGITSGAEPVHPNTLSRCNRFAPYNFEDMIRPSYGLAEAT LYVASRNSGDKPEVYFEPDKLSTGSANRCEPKTGTPLLGYGMPTSPTRVLDVDTICI ECPAGTIGELVWGDVNAEGYWNKPDETRTEGAMLVHPSAGTPDGSLRLTGDLGFLS EDEMFIYGRKMDLIVYGRNHYPEDISTEQEITGGVVAISVPVDHTEKLVYVIELK LLGDSAGEANDELVDKNNVTAALSRSHGLNVADLVLPFGSPITPTTSGKIRRAACVE QYRLQOFTRLDG" | proline dehydrogenase, similar to part of eg. TR:Q52711 (EMBL:X78346) proline dehydrogenase from Rhodobacter capsulatus (1127 aa), fasta scores; opt: 194 z-score: 264.3 E(): 1.5e-07,31.2%identity in 349 aa overlap. Also similar to all of twoB. subtilis proteins TR:E1184363 (EMBL:Z99120) (302 aa),fastascores; opt: 509 z-score: 919.0 E(): 0, 37.1% identityin313 aa overlap; and TR:E1182272 (EMBL:Z99105) (303 aa),fastascores; opt: 513 z-score: 783.3 E(): 0, 32.5% identityin311 aa overlapscore is 0.889" /codon_start=1 /transl_table=11 /product="hypothetical protein Rv1188" /protein_id="CAAL5865.1" /db_xref="GI:2695972" /db_xref="SPTREMBL:O50444" /translation="MAGFAHTLRPAMLAGRSRLGRIVERSPLTRGVVRRFVPGDT LDDVVDIVTALRDSGRYLSIDYLGENTVDADDAADAAAAYRAYLLDVLGRRGDIACDGV RPLEVSLKLSALGQALDQDKIALDNARICERARVGAWTVDAEDHTTDTLSLI SGDLURDFPWLGTGVVQAYLRRTLADCAELAAAGARVLCKGAYDEPASVAYRDAAOVT DSYLRCLURLTAGRGYPMVATHDPVIIAAVPGITRESGRSQGDFEYQMLYGVRRDEQR RLTCAGNHVRVYVPFGTRWYGYFLRLRAERPANLAFALTRDRRRARGCAER" 36010..36882 /gene="sigI" 36010..36882 /gene="sigI" /note="Rv1189, (MTV005.25-MTCI364.01), len: 290. Probablesigma factor, similar to eg. TR:O05767 (EMBL:U87307) extracytoplasmic function alternative sigma factor (sigE) fromM. smegmatis (204 aa), fasta scores; opt: 239 z-score: 301.4 E(): 1.3e-09, 32.9% identity in 167 aa overlap." /codon_start=1 /transl_table=11 /product="sigI" /protein_id="CAAL5866.1" /db_xref="GI:3261494" /db_xref="SPTREMBL:O50445" /translation="MSQHDPVSAARAHRAAYLVDLAFRMVGDIGVARDMVQEAFLRL RAPVDIDDERGLWLVVTSRLCLDHIKSASTRRERODIAAHWDGASVSSVDPADRV TLDDDEVRLALLIMLERGPAERVVFLHEIFGLPYQOINTTIGSOASTCQLAHARRR KINESRTAASVEPAQHRVTVTRAFIEACNSGDLDTLLEVLDPGVAGEIDARKGVVVVGA DRYGPITLRRHSPATVVLVAPQVCGQPAVLAFVNRLAGVLALSIEAGKITKHVLVO PSTLDPLRAELGGG" 36664..37840 /organism="Mycobacterium tuberculosis H37Rv" /strain="H37Rv" /db_xref="taxon:83332" /clone="t364" 36898..37776 /gene="Rv1190" 36898..37776 /gene="Rv1190" /note="Rv1190, (MTCI364.02), len: 292. Similar to YODA-MYCTU Q50600 hypothetical 32.2 kd protein cytail.10 (286 aa),fasta scores; opt: 331, E(): 1.4e-15, (29.0% identity in 272aa overlap), also YU14-MYCTU Q50670 putative haloalkane dehalogenase (300 aa), fasta scores; opt: 239, E(): 2.2e-09,(29.9%identity in 298 aa overlap)" /codon_start=1 /transl_table=11 /product="hypothetical protein Rv1190" /protein_id="CAAL5867.1" /db_xref="GI:3261495" /db_xref="SPTREMBL:O86348" /translation="WTMKSLAADRPSWLSSSAWHPQVLLSHHGGIAVTIDIGDPA VLVFGVHGSFVWRDVLRLANDFRCAIDAPCCGLSDRLSTPTTLQAADAITSVID ALQJDLTLVADHDLGGPAGFLAAARRGDVAALAAVNCFAWRPTGFLFRCMLAAGSA PVRELDAAINALARATSTRFGRHWSRADRAFRAGIDAPARAHAYFRDARRAHA LYTDVDAALRGGLADRPLLTIFCQFENDPLRFQPRWKELEFPTARQLQVRGNHIFPMCD PDLVAGALTSFVORST" | 6327 a 11950 c 12672 g 6891 t BASE COUNT ORIGIN | source gene CDS | /codon_start=1 /transl_table=11 /product="hypothetical protein Rv1186c" /protein_id="CAAL5864.1" /db_xref="GI:2695971" /db_xref="SPTREMBL:O50443" /translation="MDAITQVVPANRPHVDYAPKSPERTRLRTELASLADHPIDILPH VIGGRHMGDGRIDVQPHRHAARGLTINATHADAAAASAKSDWAALPDE RAVELRAADLLAGPWREKIAAATMLGQSKSVYQAEIDAVCELIDFWRNFAFASITOL EQOISGPGENRIRIDPLDGFVYAITPFNFSTIAGNLPTAPALMGNTVIWKPSIQT LAAVLTMLLEAAGLPPGIVNLVTDGFAVSADVALDPLAGIHFTGSTATFGHLNQQ VCTNIGRHSYPLVGETGCKDFVVAHASRDPVALDRAKFDYQGQKSAVSRAF IAHSVQRMGDELLAKAELRGDITDLSNYSYGNALIDQAFKVDNDAIERAKGAAT VAVGEYDSEGYFVRPTVLLSDPTDESFIYEYFGLLSVHVYDERYEQLDVIDT GSRYALTGAIVADRDQAVLTALDRLFRFAAGNFVNDKPTGAVVGVQPFQGGARGSGTND KAGSPNLRLRWTSAHSIKETFAATDHIYHPMADV" 34939..35928 /gene="Rv1188" 34939..35928 /gene="Rv1188" /note="Rv1188, (MTV005.24), putA, len: 329. Possible |
|------|-----|---|---|---|---|-----------------------|--|

| | | | |
|--------------|---|------|---|
| misc_feature | 18147..18197 | RBS | ESVCTAAGARFIGGVACCGLAELHGTGTTTYYGLTPRDRTRTPADAMTQGMFTGLIP |
| | /gene="pks3" | | ITVPIAGSAFGDAARAASTFSDGSKVLAEPYDVRVVELSSTLTMPRPFPVNFIDAG |
| | /note="P800606 Beta-ketoacyl synthases active site" | | RAPLSVLTAEITGNTNIGVSDGRYSYQLSIIVIRVEQGTAVAVWFDPDPIARESVAR |
| | 19152..23900 | | YLATLKSVFQRAVESGQQQNV" |
| | /gene="pks4" | | 25424..25428 |
| gene | 19152..23900 | gene | /note="possible RBS upstream of Rv1183" |
| | /gene="pks4" | | 25438..28446 |
| | /note="pks4" | | /gene="mmpL10" |
| | 19152..23900 | | 25438..28446 |
| | /gene="pks4" | | /note="mmpL10" |
| CDS | 19152..23900 | CDS | /note="Rv1181, (MTV005.17), len: 1582. Probable |
| | /gene="pks4" | | polyketidesynthase, similar to many eg. MCAS_MYCBO_Q02251 |
| | /note="Rv1181, (MTV005.17), len: 1582. Probable | | mycocerosicacid synthase from mycobacterium Bovis (2110 |
| | 19152..23900 | | aa), fasta scores: opt: 3518 z-score: 6207.2 E(): 0, 59.7% |
| | /gene="pks4" | | identity in1614 aaoverlap. Note that this similarity |
| misc_feature | 18147..18197 | RBS | extends upstreamofthe first initiation codon into the |
| | /gene="pks3" | | upstream CDS MTV005.16;however the stop codon at the end |
| | /note="P800606 Beta-ketoacyl synthases active site" | | of MTV005.16 is presentin atleast 4 independent clones |
| | 19152..23900 | | (BAC, cosmid and pUC) fromthe genome. The two CDS's may |
| | /gene="pks4" | | represent separate modulesofthepolyketide synthase." |
| gene | 19152..23900 | gene | /codon_start=1 |
| | /gene="pks4" | | /transl_table=1 |
| | /note="pks4" | | /product="pks4" |
| | 19152..23900 | | /protein_id="CAAL5858.1" |
| | /gene="pks4" | | /db_xref="GI:2695965" |
| CDS | 19152..23900 | CDS | /db_xref="GI:2695965" |
| | /note="pks4" | | /db_xref="SPTREMBL:O50437" |
| | /gene="pks4" | | /translation="MTASSFDELSAALRDVAGDQIPYQPAVGHDDRGPVWFSGQSQ |
| | 19152..23900 | | WPGMGTLLVAPVFAATVAAMEPVIAIESGFSVTEAMSAPOTVSGIDRVQPTIFAVQ |
| | /gene="pks4" | | VALAAALKSYGVRPGCAIIGHISGEAAAAVAGALSILHDGLRVICRSLRMSRIAGSGA |
| misc_feature | 18147..18197 | RBS | MASVELPGQOVLSELAIRGISDVVLSVASPTSTVGGATQSRDLRAWEOQDVLAR |
| | /gene="pks3" | | EVAVDASHTPQVDPIFIDELLEVLAEVDPTAPEIPIYSATLMDPRERFSFGEYWEEN |
| | /note="P800606 Beta-ketoacyl synthases active site" | | LRYTVFAAAVQAALKDGYRVFGELEAPHPLLTVAVEQNAASLDMPATLAAMRGEQL |
| | 19152..23900 | | PFGLGFAVHLLLEPERHVMQAGVTGAHPMLGDHRIHNVAAPGAAYCEMALAAAR |
| | /gene="pks4" | | AVHPLGLAHVRLLEPERHVMQAGVTGAHPMLGDHRIHNVAAPGAAYCEMALAAAR |
| gene | 19152..23900 | gene | TTGLSELVRDIFKFOETLLDQTVVSSAATIAAPGILQFAVESHOEGEPARRASML |
| | /gene="pks4" | | HALEMPQPGVDYTNALTAAHSSMGSELKMFNSLGIOYGAPSGVLAVHTARGDV |
| | /note="pks4" | | TTVLEVALPGAIRSQQSNAYASHPALLDACFQSVLVHPVEQVATVGGMLPVGRRLR |
| | 19152..23900 | | NYHSTRVYCLARVSSRAGEADLDVFDQAGTVLLTVLEGLRAGISHERANR |
| | /gene="pks4" | | VFDELLTIEWERGELPEVPQIDAGSWLLLSASEADPLTAQLADALNAVAGQSTVAS |
| CDS | 19152..23900 | CDS | ASDVAQLRSLGLRLTGTVVVTGPPTGGITQCGRDYVSQLVIGIAIELAELEPPEPLF |
| | /gene="pks4" | | VYTRSAASVPSDLANLEOAGLRMLRVIDSHPHLGATAIIDVNDIETVAALVASQLQ |
| | /note="pks4" | | SGSQDETAWRNGIWTARLPGLRPAERTAVVEYBRDGMRLQIRTPGDLSELEFV |
| | 19152..23900 | | TFDRVAPGGELEAVATASSNFADVLVAFGRYPFPEGYRQQLGIDFAGYVTAVGPDV |
| | /gene="pks4" | | TEHRIGDHVGMSSANGCWSFVRCARLAVTLPELPVAAAAAAPTASATAWIAYLHDL |
| misc_feature | 18147..18197 | RBS | ARICSDDDLTHSGTGGVQAAIAIARAAGCEIFATAGSAQRQLLHDHMGVEHYVDSR |
| | /gene="pks3" | | STFEARQLRGTDGVDVNLSPGAAQAGIELLAFGEFVEIGKRDIIYDTRLGL |
| | /note="P800606 Beta-ketoacyl synthases active site" | | PPFRNLSIYAVDALLTHSHIPTRVRLTKTYOHTVEGTLVPVOTTHPIHDAAVAI |
| | 19152..23900 | | RLVGCAGHTKSVLDVPRTGECAVVPPEOVRTSRPDGAYLVTTGGLGGLFLAGELA |
| | /gene="pks4" | | RAGCGRVLNSRSTPSPHATRVIERLRAGADIQVECGDIADAATAHRVAVATASGL |
| gene | 19152..23900 | gene | PVRVGLHAAVVEDATLANVIERLCRWAPKVVHAWNHRATAAQPLEWFCFLFSSAA |
| | /gene="pks4" | | ALVGSFGQCAANAANSWLDFAHWRRAQGLPATSTAWGAAEIGRATALAEGTGAIA |
| | /note="pks4" | | PAEGARAFQTLIRYGRAYSGYAPIMGTPWLTAFAQSRFAEAFHATGQNPATGKFLA |
| | 19152..23900 | | ELGSLPREWPTVRRLVSDQISLLRRTIDPRLPSLDYGLDSLGNLELRTRIETEG |
| | /gene="pks4" | | IRVSPKITTVRGLAEHVCDLAAQSAVP" |
| CDS | 19152..23900 | CDS | 23945..23950 |
| | /gene="pks4" | | /note="possible RBS upstream of Rv1182" |
| | /note="pks4" | | 23953..25371 |
| | 19152..23900 | | /gene="papa3" |
| | /gene="pks4" | | 23953..25371 |
| misc_feature | 18147..18197 | RBS | /gene="papa3" |
| | /gene="pks3" | | /note="Rv1182, (MTV005.18), len: 472. Similar to other |
| | /note="P800606 Beta-ketoacyl synthases active site" | | Mycobacterial hypothetical proteins eg. TR:Q49618 |
| | 19152..23900 | | (EMBL:U00010)B1170_C1_180 from M. leprae (471 aa), fasta |
| | /gene="pks4" | | scores: opt: 2526 z-score: 3476.5 E(): 0, 75.6% identity |
| gene | 19152..23900 | gene | in 471 aa overlap" |
| | /gene="pks4" | | /codon_start=1 |
| | /note="pks4" | | /transl_table=1 |
| | 19152..23900 | | /product="papa3" |
| | /gene="pks4" | | /protein_id="CAAL5859.1" |
| CDS | 19152..23900 | CDS | /db_xref="GI:2695966" |
| | /note="pks4" | | /db_xref="SPTREMBL:O50438" |
| | /gene="pks4" | | /translation="MLRVGLTIGTLLDWPSTGSTVSWRPSAVANTKASQAPISDVP |
| | 19152..23900 | | YSYMQAQRHCYCEOKAGLDIYSLRWVSCQPGQCDIRAAANYVNAHLRHRDITRSW |
| | /gene="pks4" | | FOYNGNGQIRRTIQDPADIEFVPHVHGELTLPQIREIVQNTPPDQWGCFFRGIVQG |
| misc_feature | 18147..18197 | RBS | CDHETFFASVDHVHDAMIVGVTLMEFHLMAALVGGHAPLELPAGSYDDFCRRQHT |
| | /gene="pks3" | | FSSTLTVESQVRAWTKFAEGTNGSFPDFPLPLGDPSPKPSADIIVTVMWLDEEQTAQF |
| | /note="P800606 Beta-ketoacyl synthases active site" | | |
| | 19152..23900 | | |
| | /gene="pks4" | | |
| gene | 19152..23900 | gene | /note="Rv1185c, (MTV005.21c), len: 578. fadd21, Probable |
| | /gene="pks4" | | acyl-coA SYNTHASE, similar to eg. TR:P71495 (EMBL:U75685) |
| | /note="pks4" | | ACYL-coA SYNTHASE from Mycobacterium bovis (582 aa), fasta |
| | 19152..23900 | | scores: opt: 2388 z-score: 2929.0 E(): 0, 61.8% identity |
| | /gene="pks4" | | in 579 aa overlap" |
| CDS | 19152..23900 | CDS | |
| | /gene="pks4" | | |
| | /note="pks4" | | |
| | 19152..23900 | | |
| | /gene="pks4" | | |
| misc_feature | 18147..18197 | RBS | |
| | /gene="pks3" | | |
| | /note="P800606 Beta-ketoacyl synthases active site" | | |
| | 19152..23900 | | |
| | /gene="pks4" | | |
| gene | 19152..23900 | gene | /note="Rv1184c, (MTV005.20c), len: 359. Similar to |
| | /gene="pks4" | | severalother Mycobacterial hypothetical proteins eg. |
| | /note="pks4" | | TR:Q49633 (EMBL:U00010) B1170_F3_112 (391 aa), fasta |
| | 19152..23900 | | scores: opt: 1422 z-score: 1199.0 E(): 0, 62.% identity in |
| | /gene="pks4" | | 338 aa overlap. Also similar toRv3822, Rv3539, Rv1430, |
| CDS | 19152..23900 | CDS | Rv0151c etc." |
| | /gene="pks4" | | /codon_start=1 |
| | /note="pks4" | | /transl_table=1 |
| | 19152..23900 | | /product="hypothetical protein Rv1184c" |
| | /gene="pks4" | | /protein_id="CAAL5861.1" |
| misc_feature | 18147..18197 | RBS | /db_xref="GI:2695968" |
| | /gene="pks3" | | /db_xref="SPTREMBL:O50440" |
| | /note="P800606 Beta-ketoacyl synthases active site" | | /translation="MKRVLAGAFVNLVGMWAGGFGTAIAASEPAYPWPAPPPSPSPV |
| | 19152..23900 | | GDASTAKVYVALGGARMPGIPWXYETNQAGSOYFPNAKHDLIDYPAGRAFSWMPMLL |
| | /gene="pks4" | | PPGSHQDNMTVGVAVKDGNTSLDNAIHGTDTPAAAVGLSQSLEIYLDQEOARLANDPT |
| gene | 19152..23900 | gene | PAPDKLQTLTFTGDPTRHAFASFLARIFPPGSHIPIPIFIETMPQQVDSQYDTHRVV |
| | /gene="pks4" | | TAYDGSFDPDRPDNLLAVANAAGIAAHTPIGTGFGDPPQNIIRITVNSRGATT |
| | /note="pks4" | | TYLVPNHILPLTLPRLYILGMSDAEVDQIDSVLQPDIDAAYARNDNMTFRPVSDVPVRG |
| | 19152..23900 | | LDPLTAPGSIIVEGARGLLGSPAFGG" |
| | /gene="pks4" | | complement(29533..29536) |
| CDS | 19152..23900 | CDS | /note="possible RBS upstream of Rv1184c" |
| | /gene="pks4" | | complement(29694..31430) |
| | /note="pks4" | | /gene="fadd21" |
| | 19152..23900 | | complement(29694..31430) |
| | /gene="pks4" | | /gene="fadd21" |
| misc_feature | 18147..18197 | RBS | /note="Rv1185c, (MTV005.21c), len: 578. fadd21, Probable |
| | /gene="pks3" | | acyl-coA SYNTHASE, similar to eg. TR:P71495 (EMBL:U75685) |
| | /note="P800606 Beta-ketoacyl synthases active site" | | ACYL-coA SYNTHASE from Mycobacterium bovis (582 aa), fasta |
| | 19152..23900 | | scores: opt: 2388 z-score: 2929.0 E(): 0, 61.8% identity |
| | /gene="pks4" | | in 579 aa overlap" |

/translation="MSFVTRPDSIGETAANLHEIGVTMSAHDHGGVTPLTINVESPAH
DIVSVTSMLFMSHGELYKATARAQARVHIESPVQTLQTSKTSYWLTELANRAGTST"
complement(4051. .4056)
/note="possible RBS upstream of Rv1169c"
4222. .5133
/gene="Rv1170"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1170"
/protein_id="CAAL5847.1"
/db_xref="GI:2695954"
/db_xref="SPTREMBL:O50426"
/translation="MSETPRLFFVHAHPDDDESLSNGATIAHYTSRGAOVHVVTCTLGE
EGEVIGDRWAOLTADHADQLGGRIGELTAALRALGVSAPIYLGCGRMRDSSMAGTGD
QRSQRFDADPRQTVGALVAILRELREHVVTYDPNGYGHDPDRVHTVTTAAVAA
AGVSGTDADHPGDPVTKFYVTVLGLSALISGARALVPDDLRLPEWHPRADEIAFGY
SDGWDIVAYEADQEARAAKVAALAAHATQVVVVGPTGRAAALSNNLALPILADEHYVLA
GSGSARDERGWTDLLAGLGTASGT"
5303. .5665
/gene="Rv1171"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1171"
/protein_id="CAAL5848.1"
/db_xref="GI:2695954"
/db_xref="SPTREMBL:O50427"
/translation="MVVLALITVGDVWSALAGALLMPWYIGSAPPPIISALISGLVNA
LWMAARWTTSSRVAALPLWALLVTAAMSFGGDDVILGGQGLLVYALVFFVAGA
VPPAWLWRRVQADGSG"
complement(5673. .6599)
/gene="PE"
complement(5673. .6599)
/gene="PE"
/note="Rv1172c, (MTV005.08c), len: 308. N-terminus is
similar to the N-terminus of members of the M.
tuberculosis PE family of proteins eg. TR:P71748
(EMBL:Z81368) MTCY253.25C (361 aa), fasta scores; opt:
483 z-score: 520.9 E(): 7.8e-22, 46.4% identity in 192 aa
overlap"
/codon_start=1
/transl_table=11
/product="PE"
/protein_id="CAAL5849.1"
/db_xref="GI:2695956"
/db_xref="SPTREMBL:O50428"
/translation="MSFVFAAPEALAAAADMAGIGSTLNANVVAAPVTTGVLAANA
DEVSTQVAALLSAHQGYQQLSRQMTAFHQFVQALRASADAYATAASAQAQWVNA
VNAPRALHGLPLISADASTGGSNALRSVQSMFLGTGSSALGGSAANAASGALQ
LOPTGGASGLSAGALLPRGAAAAAALPALAAESIGNAIKNLVNAVEPWQYGFNLT
AWAGWLPYIGILAPQINFYFLGEPVQAVLFNAIDFVDGTVTFESQALTNIETATA
SINQFINTEINWIRGLPLPSPGPGPSLP"
complement(6610. .6614)
/note="possible RBS upstream of Rv1172c"
6836. .6839
/note="possible RBS upstream of Rv1173"
6849. .9419
/gene="Rv1173"
/codon_start=1
/transl_table=11
/product="PE"
/protein_id="CAAL5850.1"
/db_xref="GI:2695957"
/db_xref="SPTREMBL:O50429"
/translation="MPQVGRKSTALPSPVPPQANASALRRVLRARDGVTLNVD
EAIAMTARGDELADLCASARVRDAGLVSAHRGPGSLRAISYRKVFIPVTRLCRDN
IRKSHKEFHIOEIVQVONFRAKEHTMAAAPPDAGIEDIYLLVAVALVILGPGNRIOAP
PMLVSGDECRALVGDDWGVSPPLTPDHVNPFRPWPALDELAAYTAGYDMVQRL
TAQRYVOAGAAMIDPRVRGHVADPADPATGLARDVNPVGMPOEDVDVLSRGRVDLG
AAIDTQGRNTAVRSDLASAFGDWESIIEOVHELAVRAPERIDTDLAALRSRERAPAG
CTDGEYLATADGPALENAALADSLRRDVGDEVTFVNNRNINTNICYGCRFCA
FAQRGDDAYSLSVGEVADRAWEAHVAGATEVCMQGGIDPELPVTGYADLVRAVRAK
VPSMHVHAFSPMEIANGVTKSLIREWILIGLEAGLDITPGTAABILDDEVFRWLTK
GKLPTSLMTEIYVTAHEVGLRSSTMMYGHVDSPRHWLAHLNVLRIQDRTGGTFEFV
PLPFVHONSPLYLAGAARPGPSHRONRAVHALARIMLHGRISHIOTSWVKLGVRRTQV
MLEGGANDLGTTMEETISRMAGSEHGSAKTVAELVAIAEGIGRPARQRTTYALLAA
"

repeat_unit
/note="62 bp direct repeat copy 1"
9475. .9536
/note="62 bp direct repeat copy 2"
9537. .9579
/note="62 bp direct repeat partial copy 3 (43/62 bp)"
complement(9587. .9919)
/gene="Rv1174c"
complement(9587. .9919)
/gene="Rv1174c"
/note="Rv1174c, (MTV005.10c), len: 110. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1174c"
/protein_id="CAAL5851.1"
/db_xref="GI:2695958"
/db_xref="SPTREMBL:O50430"
/translation="MRLSELTSALSGAVGAMSLTYGAGVASADPVDVAVNTTCYGOV
VAALNATDGRAAQFNASPVQSYLRNFLAAPPORAAQAOLQAVPGAQTIGLVES
VAGSCNNY"
complement(9929. .9932)
/note="possible RBS upstream of Rv1174c"
complement(10120. .12144)
/gene="fadH"
complement(10120. .12144)
/gene="fadH"
/note="Rv1175c, (MTV005.11c), fadH, len: 674; Probable
2,4-dienoyl-coA reductase, highly similar to E. coli
TR:G2584857 (EMBL:U93405) 2,4-DIENOYL-COA REDUCTASE
(FADH) (672 aa), fasta scores; opt:2344 z-score: 2419.4
E(): 0, 53.1% identity in 671 aa overlap"
/codon_start=1
/transl_table=11
/product="fadH"
/protein_id="CAAL5852.1"
/db_xref="GI:2695959"
/db_xref="SPTREMBL:O50431"
/translation="MTNPYPNLLSPDLGFTLLRNVRVMSMHTGLEPRARHIDRLAD
YFAERAGGVGLIITGGYAPNRTWLLPFASELVTSQAARRHRTTRAVHDSGAKILL
QILHAGRYAYHPLAVSASP IKAPITPPFRPALSRGVEATADFCACQALARDAGYDG
VEIMSGEYLLNOFLAPTNKRTDSMGCTPANRRFPVEIIPRSRAVNGCDFIICVRL
SMADYVAGSQSWDEIVALATEVEGACATIIISGFCGHEARVPTIYTVSPGGAFVDISS
AVAEHTIPVVASNRINMPQAAERILAEITQVRLISHEARMLUSDPMWLKAQSNRVDEI
NTCISCNACLDHAFATVSCLLNPRAGRTQVLVSPTRRSRVAVVAGAPAGLATA
ANAAQRHRTLFEANDFIGGQFDMARRIPGKEEFSETIRYFSTILAKHGVGERLGR

mj0446 from methanococcus jannaschii (361 aa), fasta
scores; opt: 801 z-score: 872.6 E(): 0, 41.2% identity in
337 aa overlap; C-terminus (aa 530-856 is similar to eg.
YE31_METJA Q58826 hypothetical protein mj1431 from
methanococcus jannaschii (359aa), fasta scores; opt: 1089
z-score: 1193.3 E(): 0, 48.7% identity in 337 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1173"
/protein_id="CAAL5850.1"
/db_xref="GI:2695957"
/db_xref="SPTREMBL:O50429"
/translation="MPQVGRKSTALPSPVPPQANASALRRVLRARDGVTLNVD
EAIAMTARGDELADLCASARVRDAGLVSAHRGPGSLRAISYRKVFIPVTRLCRDN
IRKSHKEFHIOEIVQVONFRAKEHTMAAAPPDAGIEDIYLLVAVALVILGPGNRIOAP
PMLVSGDECRALVGDDWGVSPPLTPDHVNPFRPWPALDELAAYTAGYDMVQRL
TAQRYVOAGAAMIDPRVRGHVADPADPATGLARDVNPVGMPOEDVDVLSRGRVDLG
AAIDTQGRNTAVRSDLASAFGDWESIIEOVHELAVRAPERIDTDLAALRSRERAPAG
CTDGEYLATADGPALENAALADSLRRDVGDEVTFVNNRNINTNICYGCRFCA
FAQRGDDAYSLSVGEVADRAWEAHVAGATEVCMQGGIDPELPVTGYADLVRAVRAK
VPSMHVHAFSPMEIANGVTKSLIREWILIGLEAGLDITPGTAABILDDEVFRWLTK
GKLPTSLMTEIYVTAHEVGLRSSTMMYGHVDSPRHWLAHLNVLRIQDRTGGTFEFV
PLPFVHONSPLYLAGAARPGPSHRONRAVHALARIMLHGRISHIOTSWVKLGVRRTQV
MLEGGANDLGTTMEETISRMAGSEHGSAKTVAELVAIAEGIGRPARQRTTYALLAA
"

repeat_unit
/note="62 bp direct repeat copy 1"
9475. .9536
/note="62 bp direct repeat copy 2"
9537. .9579
/note="62 bp direct repeat partial copy 3 (43/62 bp)"
complement(9587. .9919)
/gene="Rv1174c"
complement(9587. .9919)
/gene="Rv1174c"
/note="Rv1174c, (MTV005.10c), len: 110. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1174c"
/protein_id="CAAL5851.1"
/db_xref="GI:2695958"
/db_xref="SPTREMBL:O50430"
/translation="MRLSELTSALSGAVGAMSLTYGAGVASADPVDVAVNTTCYGOV
VAALNATDGRAAQFNASPVQSYLRNFLAAPPORAAQAOLQAVPGAQTIGLVES
VAGSCNNY"
complement(9929. .9932)
/note="possible RBS upstream of Rv1174c"
complement(10120. .12144)
/gene="fadH"
complement(10120. .12144)
/gene="fadH"
/note="Rv1175c, (MTV005.11c), fadH, len: 674; Probable
2,4-dienoyl-coA reductase, highly similar to E. coli
TR:G2584857 (EMBL:U93405) 2,4-DIENOYL-COA REDUCTASE
(FADH) (672 aa), fasta scores; opt:2344 z-score: 2419.4
E(): 0, 53.1% identity in 671 aa overlap"
/codon_start=1
/transl_table=11
/product="fadH"
/protein_id="CAAL5852.1"
/db_xref="GI:2695959"
/db_xref="SPTREMBL:O50431"
/translation="MTNPYPNLLSPDLGFTLLRNVRVMSMHTGLEPRARHIDRLAD
YFAERAGGVGLIITGGYAPNRTWLLPFASELVTSQAARRHRTTRAVHDSGAKILL
QILHAGRYAYHPLAVSASP IKAPITPPFRPALSRGVEATADFCACQALARDAGYDG
VEIMSGEYLLNOFLAPTNKRTDSMGCTPANRRFPVEIIPRSRAVNGCDFIICVRL
SMADYVAGSQSWDEIVALATEVEGACATIIISGFCGHEARVPTIYTVSPGGAFVDISS
AVAEHTIPVVASNRINMPQAAERILAEITQVRLISHEARMLUSDPMWLKAQSNRVDEI
NTCISCNACLDHAFATVSCLLNPRAGRTQVLVSPTRRSRVAVVAGAPAGLATA
ANAAQRHRTLFEANDFIGGQFDMARRIPGKEEFSETIRYFSTILAKHGVGERLGR

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 37840)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
Tekaiia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jageis,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Sqaures,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)
2 (bases 1 to 37840)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2635948.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
FEATURES
source
1. 37840
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..36758
/note="fragment designated v005. Does not represent a
physical clone"
70..1377
/gene="lpqW"
70..1377
/gene="lpqW"
/note="Rv1166, (MTV005_02), len: 670. Possible membrane
protein, almost identical in part to M. tuberculosis gene
fragment TR:G2384665 (EMBL:AF009358) ORFA2-898 (FRAGMENT)
(59aa)(93.9% identity in 49 aa overlap) (Nano F.E., Doran
J.L., Treit J.D., Moran A.J.; Identification of
Mycobacterium tuberculosis peptides that stimulate immune
human peripheral blood monocytes; Unpublished). Contains
possible N-terminal signal sequence and appropriately
positioned PS00013 prokaryotic membrane lipoprotein lipid
attachment site. Also contains"
/codon_start=1
/transl_table=11
/product="lpqW"
/protein_id="CAAI5843.1"
/db_xref="GI:2695950"
/db_xref="SPTREMBL:O50422"
/translation="MGVPSVRRVCVTVGALVALACMVLACGTVSPPPAPQSTDPTRS
TPPPRRPTQIICIDWIGFGENPHLLSDLSVNAALISALVPSAFRPIPDNPTGCS
RWEMDPTLLVSADVTNNHPFTVYKIRPEAQWTDNFIADDDFWYLWQQWVTPQGVVD

PAGYHLITSVQSLGEGKQAVVTFAQYPAWRELFDTILPAHIVKIDPGCFASGLARAL
PVTGQFRVENIDPQRDEILLARNDRYWGPPSKPGIILFRRAGAPAAALADSVRNGDTQ
VAQHGSGAAFAQLSAIPDVRTARIVTPRVQOFTLRANVPKILADQVRKAILGLLDVD
LLAAGACTDNTVLDQAIRSPSDGVVPTAPPAMSSAAALGLEASGFQVDWTSV
SPASVPDSTTSVSTGPGVEVIRGSKDQGLTLVIGVAANDPFSVAVANTAAQBLR
DVGIAATVLAIDPVTLYHDAINDNRVDAIVGWRQAGNLAATLLASRGCPALQATTVR
AANAPTAPSAPIGPTPSAAPDPTATPPTAPRPSDGCALVKAFPSNLTCIDCRSTQSN
IDAALNGTKNINDVITAVEPRLWNNSTVLPILQDPTTIVAAGPSVQNSLSGAVPVGIV
GDAGQWVKTGQ"
121..153
/gene="lpqW"
/note="PS00013 prokaryotic membrane lipoprotein lipid
attachment site"
complement(2005..2610)
/gene="Rv1167c"
complement(2005..2610)
/note="Rv1167c"
/note="Rv1167c, (MTV005_03c), len: 301. Probable
transcriptional regulator, similar to several eg.
TR:D1022772 (EMBL:D85417) hemR from Propionibacterium
freudenreichii (243 aa), fasta scores; Opt: 268 z-score:
416.1 E(): 5.4e-16, 35.9%identity in 198 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1167c"
/protein_id="CAAI5844.1"
/db_xref="GI:2695951"
/db_xref="SPTREMBL:O50423"
/translation="MTVSAPAKANPYRRRGEVLERALYDATLAELESAGYGLTMEGI
AARQTKAALYRWAGKRELYLAOVQALPVPPPEPRADRSARENLLAVFTANCEILA
KTAIPSMIEIVSQLLHEPRLAIFINVSYPALRIVESILOAGVRSGEIDPATLTPMT
ARIGPALIHQVLFTHGSPDREQLTRIIDAMILTTGERRES"
complement(2682..3722)
/gene="ppe"
complement(2682..3722)
/gene="ppe"
/note="Rv1168c, (MTV005_04c), len: 346. Member of the M.
tuberculosis ppe-family, similar to many eg. TR:E332789
(EMBL:Z98268) MTC125.27C (385 aa), fasta scores; Opt: 504
z-score: 780.2 E(): 0, 36.6% identity in 388 aa overlap"
/codon_start=1
/transl_table=11
/product="ppe"
/protein_id="CAAI5845.1"
/db_xref="GI:2695952"
/db_xref="SPTREMBL:O50424"
/translation="MDFTIFPPEFNSLIQGSARPELVAAWKNLSNELSYAASRFE
SEINGLITSWRGSPSTMAAVALPRAVIVTASLAELVADHSIVSACAYEAAHRAHV
PLPVETNRLRLALATTNIEGHTTFAIFALDIAQYVSDQSGAMNDYATMAAAAR
LTPFSPAPIANPGALARKLYELIGSVSETVSFAAPATKNLPKLTLLTGTGYPLTA
ARISSIPVEYVLAIFVEGSMQGMNLAHRSITPTLKGPLELLPNAPVAVSATLGNA
DTIGLSVPPSWADKSIITPLAKAVPTSPAGPGSGTSAQLGLASLAGGAVGAVAART
RSGVILRSPAAG"
complement(3740..4042)
/gene="pe"
complement(3740..4042)
/gene="pe"
/note="Rv1169c, (MTV005_05c), len: 100. similar to the
N-terminal region of the M. tuberculosis PE family of
proteins eg. TR:O05297 (EMBL:Z93777) MTC1364.07 (99 aa),
fasta scores; Opt: 209 z-score: 407.6 E(): 1.6e-15, 37.4%
identity in 99 aa overlap. Also similar to the N-terminus of
M. tuberculosis TR:P77909 (EMBL:U76006) ESTERASE/LIPASE
(EC 3.1.1.3) (TRIACYLGLYCEROL LIPASE) (437 aa), fasta
scores; Opt: 193z-score: 381.8 E(): 4.4e-14, 37.2%
identity in 94 aa overlap. Contains a helix-turn-heix
motif from aa 88-109(+2.76 SD)"
/codon_start=1
/transl_table=11
/product="pe"
/protein_id="CAAI5846.1"
/db_xref="GI:2695953"
/db_xref="SPTREMBL:O50425"

```
...
258228 CTTGCATTAGATTATGCCCTTTCATAGGCTTAICTATATAAAGAGTATGAAGCATGGCA
      |-----|
      258278
258288 CATCAAAGCCAGTTAGCCACATAATCCACCACATAAAGTAGTTTAAATTCATCGCTATAT
      |-----|
      258337
258348 TTCTCATACGATTGGCCAAAACGCTTTTGAATCTTACTGTGAAGTGTACTAAACT
258408 CTACCCCATCGCTAGAGCTTGAAGTCATCACTACTTTTATAGCCCGCTACTTAGCTCAT
      |-----|
      258415
      258466
---|
258468 CGCTATGCCCACTTGGTCTTATTTTATGATTTTCATCATATAGTTTATAGCTA
      258471
...
258551 TTTCCGCCGAGCTTACTAGACCTTGTTCOAAGTGTTTTACTATATCTTTAGCGATTGT
      |-----|
      258601
258611 TTTAGTCTGCTTGTGCTACCTACTAAAGCTTCTAGCTTAGTACTT
...
258810 CTATGTAGTTTCCAAAACCTGCAGGGGTGTATATCATGCTTTCATTTGGGCTTCCTG
      |-----|
      258860
258870 TAAATCCTATATAAGTAGCATTTGGCAAAGCATCACGAGATATTTTGCGAACCATATAA
258930 CTATCTTTTGGCCCTAGCACATCGCCACCTTTTGCATTTTCTTCTACTACTTTTGCATCAA
      |-----|
      258949
258990 AGCGTATTGGGTTCTGTGTCTCACTGCAATTAGCATATATATCTCTAAGCGATA
      |-----|
      259043
259050 AGCACTCAATACATATATCATTTTCTGGGACTAAATTTTGAATAGTAGT
...
259100 AATACCACTCCACCCTAGCTACCCCTTAAAGCTCTTTTAAATGCTCTTTATCGTTGCT
      |-----|
      259150
259160 TGTTTTGGCTCTGTCTCTTAAAGGTTTCGCACAACCTTGCAAAAGTAC
...
259224 GATCATTTCTATCCGTGATAAGCAAAATCGTAGGATTTTAAAGCCTTTATCGCTTAG
      |-----|
      259274
259284 CACTATAAACACCATAGACAAAGACTTCCCCTACTTCCGATG
...
259373 GTGATTGTTCTACTAATGGCTTTATTTACTGCGTAGTATTGATGATGAAGGATTTT
      |-----|
      259423
259433 TTGGTAGTTTTTATACCTTAAATACCATTAAATCTCTTATTGCT
...

```

```
259541 TCTATTAAAGTTTTAATTTCTAAAGACTTGTTTGGAGCTAGTTTATCCCATCGCTACTT
      |-----|
      259591
259601 TTCCACCCCATAAACCTTGAAAAATCAGCGAAATACTCCCGCCCTAGCAAAATACCCA
      |
      259660
259661 TCGCTTATCACACAAAAAGCATTATATGTAAAAATACTTGGGATTTGTTGTTTATAAGTG
      ----|
      259665
259721 TTTATTGTGATTATAAGCTGAGTTTATCGTGGCGTTTTCATCGCTTGGGCTTTTATAGCTCC
      |-----|
      259750
      259759
259781 ATTAACAAGCGGAATTCATTTATATATAGCA
...
260644 GGATATAAATTATCTTTTGATATTTTATTGTATTGTTTCTACCTTCTTCATCGTTATCA
      |-----|
      260694
260704 AAAATACATATAAAATATGTTTTAAAAGAAGTATAACTGCACGTAT
...
261259 ACTATTCTTTGAGCACCACTCCCTTTATGATCCATTAAAGTGGTCATTTCATCGCTAAA
      |-----|
      261309
261319 GTAACAGATGTCATAGTTTTTACAACATCTCTTAATTTTTCAAAT
...
261499 CTTAAAGGCTCACTTTGCTTTTTTTCGTTATTATCTAAGGCTAACAAAGCATCGTTTCT
      |-----|
      261549
261559 AAAATAGCAGAAATTAGCTAGGCATATTCACATTACTTGCTTCTA
...
262295 TTTACCATCTCTACTCGCTGATTAAACCAATAATTTTATTATTGTACTTGAAGCATACCC
      |-----|
      262345
262355 ATTTTACCTATCTCAGCACCAGTCATTGCAATAAAATAATATCTCCAG
...
263226 TTGTTTAAAGCTTGGTTTAAAGATAATGACGAAAGCATGACAGAAGAAAAAGCTAAAA
      |-----|
      263276
263286 GCTCGTATTTTAAAGCTATGCTGATATATAATTTAAGCTATTTTAA
...
-----
644 matches found in sequence:
mtv005 ; TOIG of: mtv005 check: 8654 from: 1 to: 37840
(from "Mycobacter_ge.seq")
TOIG of: mtv005 check: 8654 from: 1 to: 37840
LOCUS MTV005 37840 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 51/162.
ACCESSION AL010186 AL123456
VERSION AL010186.1 GI:3261493
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis H37Rv

```


253043 AAAAAGGCATAGAAACCTTTTAAATACTACTTATAAAGTAGGTTTAAAAAGCGATAGAA
|-----|
253093

253103 TGGCAATTATGCAGAACTCTAGTGAAGATATAGAGCATAAAAACTC

...

253531 CTACAAATTAAGCACTAGCAGGAGTGATTAACACAAAGAATTTCGCCCTTAACGCTTCCT
|-----|
253581

253591 AAAATGTGAATCCCATATAATAATAACTACTAGTTCCTATATAATAGTTG

...

253712 ACAATATCCAAGAAACAACTTAAATAGAAATGAGATTAAATAAAGTTAATCGCCAAAG
|-----|
253762

253772 GGTGAGCTCCTTTGGCTTTTATGATTAAACACTGCCATCAAGGCCCA

...

253928 TTTCACCTCTTTGTGATCTTTGTGAGTTGATTTTTTGTGTTTTTGAATAAATATCGCTATCA
|-----|
253978

253988 TAAATATCAANTAGAAAAAANTAGGAATCATAGTGGTTACTAAAC

...

254151 AATTTTTTAAGAAGTTTAAACAAAAAATGAACCAATACARCAAGAAAAACGCCAAAA
|-----|
254201

254211 GGTAAATAAGAAGTATCCGTATATAATTTGCTAAAAATCTTAGATGTT

...

255121 TATTTTATCTGCTTATTTTCGTATGAATCGTGGGTTATATGAACCTCAACGCTTTTT
|-----|
255171

255181 AAGCTCCATATCTTTTAAACTTTTAAACCAATACCTTAAGAGAACCTAGAAACGATGGTT
|-----|
255230

255241 AATTATCAGAAAGAATAACGCTCAAAATCCACCCCAAGGTAGAATA

...

255527 ATTTTTCAGTAAAAAATGAATTTCAAGTATAAAGGGTAAGCCAAAAATTCGTTTCAA
|-----|
255577

255587 ATAGGCAATTAGCAGTGAGATTAAACATCTTTTATGAACCTTTTAA

...

255716 TTTTAGCAAGAAATATAATAGATGAAAAAATAATAAAGAGCTTTTAAACATCGCTAAGC
|-----|
255766

255776 AAAGGGATTGACATTAACGATGGTTTTTGTGATAAGTATTAATAATATCTTTTGTGAT
|-----|
255793

255836 ATCAATGACATAG

...

256024 AGGTTATCAGCTGATACTTGGAGCAACTTCGTAGGTGGCAACCCACTATCGGCATTT
|-----|
256074

256084 CACCCCAAATTTAAGCAAAATCCAATAAATAGCAGGATTATAAGCGGTTTGAATGGTTG
|-----|
256128

256144 AGATGATTAAATCAAAACGTTCTCTTTTACAGCGTTTTTTATCGCAGCTTGTGTAATAATGAGC
|-----|
256159

256204 TTACACCCATAGCCAAAGCATCGGCTTTTTTGTGTTTTCATTTCTAGCAAAAACGCTAAAGCT
|-----|
256253

256264 TTGCACCCATTTTACAGCATATTTAACCGCCATCATACCAAGTC

...

256514 ACTATTGTAAGTAAAGATGGTTTTTGCCATTTTTCGCAAACTGCTCTTGTGAGCGTTTGCA
|-----|
256564

256574 AGCATCGCACTACCGCATGAATTTACCATACAAACCCCTGCGTAATCTCCTACTTT

256634 AAATTTACTCACATTTTACCACAGCGATAACTTCTCCTGCTATCTCATGCCAGGCAC
|-----|
256658

256694 GCAGGATAAGTAGCTTCTCCCACTCGCTTCTGCGGTGTGAATATCGCTATGACAAAT
|-----|
256739

256754 GCCTCCATATAGGATTTTAATCAAAATATCATTTATCGCCATAGCATGGCGCTAAATTC
|-----|
256787

256814 AAAGGTGTAAATTTTCGCATCCTTGCTAA

...

256998 TAGTGAATGTATAAAAATTAAGCAATTTTATAACAAAGACATAGTTGCAACGCAAAAT
|-----|
257048

257058 CAAGCTATCTATAATAAGATAAAAGAAAAAATAACAAACAAATTAATTTTCAGTATCGC
|-----|
257113

257118 TTAATTTCTACTCGTTAGCTAATTCCTTAGCCAAAAATTTTCAGCTTGAGCTA
|-----|
257118

...

257792 TCATTTTCTTTTAGCAAACTTTTAAAAACCTTAGCCTTTATCGCATACATAAAACGCTTTTA
|-----|
257842

257852 TTCTTTTATGCTCAAAAGCTTTTTCGTGTGGCATAGCTATCGTGTAGG

...

258029 TTTGGCAGCATTTGTTCTACGATTTCAGCTTTTCTATCATCATCTTTTACAGCGTCTTCT
|-----|
258079

258089 TGAGTAGTGATTATATCGCCCTTGCCACCACCTTTTCACATATAAAATGATAAAGCTTTTTC
|-----|
258103

258149 AAATCACTTG

246511 ATTCAGCGTTCCTATACGATGCTCTGCAACATTAACCAATCATCTGTTCTTCTCCTGCG
|-----|
246571 ATGGTAATATAACCAATTTTATCATATAAAAGCACCATCGCCACTTAAATATATACAGCTTTT
246606
246631 CCATCTTTTTAGCTTGAGAAAATAACTTTCTATATAGCGTTTCATCATTTGCCCAAAAT
|-----|
246668
246691 CCACGAATCATCGAAGGCCAAGGTTTAGTTATA
...
246781 GCAAAATCCCTGGCAAGGCAAGGTGGCACAACTGGTTTAAAGGAGTGGCGCCTGGT
|-----|
246831
246841 AAAGCGGTGATCATATATGCCACCTGTTTCTGTTTGCCACCAAGTAT
...
247044 TGGCAAGGTGAGAATTTGCTTAATTTGATACTCTTCTATCATCTCTCCACCACGCCAAGA
|-----|
247094
247104 ATTTGGATAAGTAGAGTTCCTTCATGCAATTATAGTCGTTGCCCCACATGCTAAAGGCC
|-----|
247138
247164 ATAACTACATAAGTATGTCCTGTAATCCA
...
247666 AACATTACTATAGCTGTTCTCTGGAATCATAGGCATATAAATTACAACCCATATCGCCTTT
|-----|
247716
247726 TTFACTCGGAATTTTTTAAGCAAAATGGCCGCCCTTGCAAGTTTCAT
...
247809 CATTTACCTTCAAAATCAAGCAGCTTTATTGGGCTTGTTTTCATATCAGCATCAAG
|-----|
247859
247869 ACATTGATACTTACATTTAAAGTCCACCTTCAACCATTTATAA
...
249365 TAGCTGTGGTTATGGTTCTAGTATCGTATTTGATATTTCCACCCTACTATCATCGCTACTC
|-----|
249415
249425 CTACTTCTGTACAAATGGGTCCATAAGCTAGTGCACGACCCCTAATCAAGCAAAAGCGTA
|-----|
249459
249462
249485 ATTCATAATCATCGTTTAAACAAGTTTTAAAGGACTTAGTGAAAAGATTTTATAAGTA
|-----|
249496
249545 AAAAAATG
...
249554 TTTTATTATATTTTCTATCAAAATTTGAAAAAGTGGGATCACTATAGGAAGCGCTAAGA
|-----|
249604

249614 TAAACTGCCCTAAAATTAAGCCTTTATAGTAAAAAGCAATCCAAA
...
249708 TACTGCAACTGTGGAAAAGAAAGGCTAGTATCAACTATAAGTTTTATAAAGCGCTTCAA
|-----|
249758
249768 TTTAAATTCAAAAAACCTAAGATAAATCCCAAGGAAACCTATA
...
250366 AAGAAAGCTTAGAACTTATCAAAATGAAAAGCTTACTTTCATCTCAAGAGCGATAAAT
|-----|
250416
250426 CAGGCACGTGACAATAAAGAAAAAGCCCTTGGAAAAATCTTTGGCGG
...
250483 AGCAAAAGCTGGTATCAACAAAGCGGACAAGGTATGTTAGCAAGCATTTAAATTCGCTGAAG
|-----|
250533
250543 AAAAAAAGGTGTGATTTTAAACCGATCGTGGTACTTATATCAAAATATGAAGCCAATGAA
|-----|
250603 AAGGCAACCAAACTTAGTCATTTGTAATGAAGCGGATGATAGTCTTAAAAATTTTTATT
|-----|
250635
250663 CTGTTATAGCGACAAATCCTAAGCATTTGTAATAATGTAATAATATACAGAAGCTAGTAAT
|-----|
250670
250723 TTA
...
250960 GTCAAATAATCTTCAAAAGTGATTTAAATAGTGATTTAGGGGAAAGTTTGGCGCTATA
|-----|
251010
251020 AAATGGGAATGATGAAGAAATTTTAAATTTGTAAGTTTCAAGTTTCAAGAAAGCTTGGCT
|-----|
251067
251080 TTCAATGCGGGTGATCCTTTGTGTGATGGATGAACCTTAAATTT
...
251256 GCCTTTGCTAAGGCAAAAGGTATGAAAATACAGCATTTTAAAGCCCATGGCGGCTTTAT
|-----|
251306
251316 AATATGCAGCTATTGATGAAAATTTAGCCCTTAGCCTTTGTGAGG
...
251469 GATCGTGCCTATAATGATGATGGCCTTTAGTTTCAAGAAAAATTAGAAGGCGCTTAATA
|-----|
251519
251529 CACGATGAAAATTTAGCTATCAAAAGCGGTGATAAAAAATGATAAAG
...
252707 AGCTTGATAATCAAAAAATCAAACTTCCAAAGTATATCAAGCAAAATAAGCGCATTTT
|-----|
252757
252767 TAGAACTTGATTTAGCAAAAAATAGGTTTTTCGTGGGTATTTATGCGT
...
252757

239177 GAACAAACACATACAAAAATAGAACTTAAATACAAAAATAAAGCGTATTAAACGCTTTAT 239227
239237 TTTTCATCTTAACAATACCTTTCATTATGCAAAATTCGCCGATAA
...
240341 TTATAGAAATAGCGAATCTTTTAGCTTTTAAAGCATTAGCGCAGACAAAAATCGCTCATA 240391
240401 AAAATTTAGATGAAAAATGCTTTTAGAAGCTTGAAAAAGATTTTTTCTCC
...
240938 AAATGCTGATAGAAAAAGATAGATAGTGTCTTTAGCTAAAGAACTTTGTAAATCGTTATGA 240988
240998 GGGTATAGGTGGGATGGCTTTATCGTAATATTGCTCATGAAAAAATACGACCTTTGAAATG
241058 GGAATTTTATATACGATGGCTCAAGGGCTGCTATGTGTGGCAATGGCTCAAGGGCTGC 241071
241118 GGCACATTT
...
241947 AAAATCCATGGGTATAGCCCAAGCTTTAGTTGAAAGTGTACAGTACCAGTCCGCTTCGC 241997
242007 AAGAGACCAACAATCTTTTTGGAGAAATGGACTTGGGGTGAGAAA
...
242100 AATTTTGTAGTCTGCAAGATAGTTGACTCTTATGTTTTTAAATTTAAATCGTCATTT 242150
242160 TGCTTATGAAAAGTTTCGCGATGCAAGAGCTAAATTTGAAAGCGAA
...
242250 TTATTCGAACGCAAGGGATATTATCAATCTTTATACAAAAATCATCAAAACGATACAA 242300
242310 CCTTGAAAAATACGATACAAAACCTCAAACTCAAAACACACCTAAAAATCCAAGT
...
243422 ATTGGGGCATAAAGCCCCCTTAAATTACATTTTGAAGCACCTTGTAAAGTAGCGCTATC 243472
243482 ATCCAAAGACTTTTTTCATATTTTTCGGATATTTTCTTGTGCAAAAG
...
244109 CAAGTTGAAGTAAAAAGTTTCATAATAATACATCAAAATATCCGCGATGCCAAGCGTCAAGT 244159
244169 TTTTCAAGTCTATACATCAACACTCAAAATTTTCAGCGATTAAGGCGTGCATAGTTGCA 244204
244229 TCACATTGAGCGTTTAAAGTTCGTGAAGATTGACTTTTATTTACTTTAAAGAATCTAAAAAT 244237

244289 TTAAATGAATATCAGGGATAATAGGTGCTTTTGTAGCGTCTTTTGTGATTTTTTAAAAAGTT 244324
244349 ACAGTTTTATTTTTTAGGTCCATCCATGATGA
...
244463 AAATTTTCAAGCAATAAGCTGGATATACCATTCATTAAAGCCTTTAAAGCATCGCTATAT 244513
244523 GAAAAAGTACATTGATATTGGCATTTTTTGTGTGTACATAATGAT
...
244654 ATAGCCACAAAAACAAAGTGGAACTATACCTACGGCTGAAAGTATGCTAAACAGCCTCC 244704
244714 TACATTTGCAGGGATAAAAAAGCATTTAATACCTAAATTTTCACCT
...
244806 TTTTAAATCTTGCATATCAAGTTTAAATGTTCAAATGAAGTTTAAACAGCATATTT 244856
244866 CTTCTATAGTACTACCTGCTTTTGTGTGATGATTAAAAACAAACTTTC
...
245009 ATATCAGCAAGGCTTTTACACCACAACTTGAACCTCCCATTTCCACCTAAACAGCATATTT 245059
245069 TTTACATACTCTTTTATCTTGAATAAATGCAAGCTCTCATCGATTA
...
245219 TTAATAAAAGCGTATTGTTTTAACATTTTACTTTTCTCGTAAATAATTTTGTGCGCTCTA 245269
245279 CAAAGCCTTCTACACTTCCACAATCAAAAGCGTTTCCCTTTGAAATTTATAAGCTAAACCA 245306
245339 TACCATTGGTTGCTTGAGTTAAA
...
245553 ATAGTTTGAACCTTGCTCTTTTGGTACTTCCATCACAGCTATGATGGTGGAGCATATTT 245603
245613 TTCATAAATTTTTACCATTGGAGCCATAACATTTAAGCCTTCTTCATTAAACACACAAATC
245673 ATCAGCCAAAATTAGGCCAAAAGCTTCATCGCCCAACCAAGGTCTACCTTTTAAACCGC 245700
245733 ATCACCCTAAGCCTTTTCATTTGATTTTGTCTTGTAAAAAGTAAAGGTACAAACGATTGATCAA 245780
245793 AGACCGAATTTTCATCAAGTAAGTATCTTTTTTTGTACCTGAA
...
246451 TCCAAAATACTCACTACAGCTGATTTCTGTACACTTGGATCTTTTAGCAATAGCGCTTCA 246501

234449 AGTGGGAACTAATGCTTATTGTGAAGCTTTCTAAAGCATGATATATATATCTTTAAA
234484
234509 AACAAATTGGCTATATATTTCATCTTGCTTAA
...
234728 TTAAGTCTTTTTCATTTGCAATTTTAAACAATAATTTTGTGATTTTGTAGCGTTTCAA
234778 |-----|
234788 ATTTATTGCTGTTTTGAACAATAACAACTAACAGTAGAGCTTGT
...
235033 GCATAAAATTTATTGTCTTTTGTGTTTAAATAATCCACTTAACACATATCGCCATTT
235083 |-----|
235093 TTAACATTTTACCATTTACTTTGCGGATTTTACCAAGCTCTAGG
...
235426 GCAATAATTACTGCATCGATGGCGTGATGAAGATGTTATTCTATCTTACGCTAAAG
235476 |-----|
235486 CCCCAAGTGTCTTTAAAGCGGATGTAAGCATACCGCTTTTGTCTTCTACATGCACTTTT
235546 GAGCCTTTTGTGATCGTTTAAATTTGTATTTTTCATCATCGCTTAAAGGTAAAAATCT
235560 |-----|
235606 AAATAATCTTTAGTGTAGTTTAAACAAGCCTTGTCTATGTAGCGAGTATCGTTTAAATTT
235653 |-----|
235666 CTATCTTTAAAGTTTTTTTGTCTCTTCTCTCTTTGTAATTTTAT
...
235996 AAAATAATTTTGTCTATTATTTTAAAGCCTAGTTTTTTCACATTCAGGTTCGGCTCTTTT
236046 |-----|
236056 TTGCGCTTGTAGTTTTCGTTTGTCTCTCTTTTCTATTTTAGCTCTTT
...
236153 CTTTACCGTATTTTAAAGTAAGCATTTAAGACCTTTCTATATTTCTTTAATCGCCCTTA
236203 |-----|
236213 AAACCACAGGATTAGTTACTTCAATCTTTGTAATAAGTTTCATTAAA
...
236569 TTTTAAATCAATAAAATAAGTTCCCTTTTCTCCCTTTAAATTCATAATCATCGCTTAAG
236619 |-----|
236629 CCCAAAAGCTTTTGTGTTTATAGTTAGAGTGCCGTTTTTTTAAACTTTCATTAAA
...
236689 AGGCGTTTAAATCATCTTTTGTATATAAAATTCCTTTCAGTATTTTAAAGATTGTTTA
236691 |-----|
...
236753 TGATAATCTTGTGAAGTGCTACAAACATAAAAGCTAAGGGCGAATTTTTTGGCGCTCTTT
236803 |-----|

236813 TTTTCATCGGTAAAAAACTACAATTTCCACCAAAATGAGAAAAATC
...
236940 TTTTAAAAATAAGCTTTAATTCATCTTTTAAAAAACTTTGCGCTATGCAAGCGTTTATA
236990 |-----|
237000 ACTTTCTTTTATTTCTAACATTTGTAATTTCTTTGGAAATTTTCT
...
237348 ATTTGCGATTAAATGTTTTTAAAGTATTTAGCTTTGCTTTGCGTCTAGCAAGCGCTTTTCT
237398 |-----|
237408 AGCACTTCTTGCAGGCTCTTTGGCAAGGCTTAAGGATTCCTCTGTTTGGGATTTTCAAC
237422 |-----|
237468 CTTTGTAAAA
...
237764 GAGAAATCATCCCAATTTTATGATATGTTCTACACTTTTAGCGGTAATTTGCTGCTGCTT
237814 |-----|
237824 ATTATATCTAAATAATAATCTCATTTTGTATTTTCTTTTCAAAAT
...
237932 AGAGCATACAAAGCTTTTAAATGTTTACTTGATAGTATCTTTTATCACAGCGATTTT
237982 |-----|
237992 AAAAGTCAGCTTTCTCGAGTAGGGTTTATTATCATGAGTGTGGCAG
...
238120 ATAAATCTTAAATTTTAAAGTGAACCTTTTGTGTTGGAATGGCTTTAAAAATCTCATTT
238170 |-----|
238180 CTAGCTATGCTGGATTGCTTTTGTAGCATGTATTTATCTCTGTTTT
...
238686 TGATTTTAAAGCTATGCTGAAATTTTATCTCAGTCTGAAGCATATATCATCGCCGCTG
238736 |-----|
238746 GAGTATTTGCTGAAGGCATTTAAAGCACTTGGTGGAGTTAATATCTT
...
238911 CCGCTTTTAAATACTTTTGGAAAACTAGCTCTTGATATAGCTACAAAATTAGCGTTGCAC
238961 |-----|
238971 CTATTACTCTAGTTTTTACCTATAGAAATTCATCTCTTTTAGGTCG
...
239064 CAAAAGTAGCTCCAATGGATATTATAAAAGCAACCACTCTCTTTTACTTATCGCTATGC
239114 |-----|
239124 TTGTAATGTACTTGTAGCCTTTTATCTAGCACAAACCAATCTCTT
...
... |-----|

227438 CGCAAGTAAGGATTTGAATTTTCAAAACCTTCTTCTATGTAAAGAAAGTCGTTGGNA
|-----|
227488

227498 TCCTATTTTAAAGACTTAAAAAAGCAAAATCATATTTCTTTAAAAAA

...

228166 GTTGTGTTTTTTCACAATGGTGTAAATTTGATCTAAAAAACAATTTTATCGCCGATT
|-----|
228216

228226 TTAAGCTTTTATCAAAAGCGGTATTTCTCCATTACATCACCCGGTA

...

229016 CCCAAGTCCGCTACAAGTACAACCTTTTAAAGTATTTCTCCAACAATTTTCATCGCTCCTG
|-----|
229066

229076 AAAAAGCAAAACCTTTTAAAGCTAAAGCACTTTTAGCACCACTTTT

...

229708 AAATTTTAGTCAAAAACAACACTCAAGAAGCTACAACCTATACATTTGGCATGGCGTTCTCG
|-----|
229758

229768 TGCCACCTGATCAAGATGGAAGCCACATGATCTCTTAGCAGGAGAGAAAGAATAT
|-----|
229828
229828 ATCGTTTTTGAATTCCTCAAGATAGCGAGGAACCTTACTGTATCATCCACATCCT

...

229939 AAGCAAAAAAAGATGCTTTAAGTCATTTTAAAGAAAGAAAGATTTTGATGATTAAGCGATTGC
|-----|
229989

229999 GTCTTGATGAACGCTCAAATTCCTCAACAAACCAACCTAAATGATTTGGCTAAATGGTAGAG
|-----|
230010

230059 AAGGAGA

...

230230 ATAAAGAAGAACTTTTCTTAAGTCTGCTTCTCGTGTAGAAGTTTAAATCGAGCTCCAA
|-----|
230280

230290 AAGATGGGAATTTAAGCTAGAAGTCTTATTATGATAGGGATAA

...

230479 ATCATATGCAATGCATGGAAATGATGGGAAAAAGTGAATGAATTAATAATCGCTCTTG
|-----|
230529

230539 CGTCTATGTTTTTAATCAATGGCAAAAGCTATGATTTTAAACGCAT

...

231211 ATTGTGCGTGAAGAAGGAGGATTTAGGCGACTTTAGTTTGTATATTATGAATCGCTTTTA
|-----|
231261

231271 AAAAAATGGTTGATGAGTGGCAAAAAAGAGTAGATTTTGTATGATTTAGCTTGCCTTTT

231331 GCTCACTCTATAGCGGATGTAGCGATACATGAAGCTCGTATTTTTCAGGAATTTTAAGC
|-----|

|-----|

231343 231351
231391 AAACAAAGAAAAATTAG

...

231800 AAAATGGCACCTTGTTAAAAACCTTTACAGGCTCACTCATGAGCGAAGAGGATACCTT
|-----|
231850

231860 TAGTCCCTGTTGAAAAATATAGCATAGAAAAATAATACTCTTTTAT

...

232100 ATCTAGGAGAAGCCTTTGAAAAATCCTGCTCAATCCGCTCTTCAATCATATCGCATAG
|-----|
232150

232160 CTAATTTAGCTAAAAATTTAAATTTCCGATACAGGGTTTTTCCCTCT

...

232361 TAAACCAGGTAGACATATCAAAATAGCTAAGGCTAATGAAAAATTCATCATCGCTTAC
|-----|
232411

232421 CAGGTTTTCCTTACTCGGCTATGGTAATGTTTAATCTTTTACACAAG

...

232514 AAGCCTTTTACAAGGAAGTTTACAAGAAAAAACACCTTATTTGGAATTTTCGCTTGC
|-----|
232564

232574 ATATAGATTTTAAAAATGGAGCGCATTTTAGCCAATCTTGAAGGCAA

...

233366 CATTATTTGAATTTTATATTTTAAATTTCTCCATGACTATCTATATCAAAAAGCGTTAAAT
|-----|
233416

233426 TGGTTTCGTTATTAACTTTTTCGTTTAAAACTATGCCACCACCAAGCAA

...

233754 GTTCATCTTCATACAATAAAGCATTTTAAAAATTTGCCACATATGATTTATCGCTCTGT
|-----|
233804

233814 TTAAGGATAATTTTGCCCATCAAAATTTGATTTTACTTTTGTAAATTTGCCAATCAACGCTC
|-----|
233867

233874 TTTTGTCTCTTCTGTTAAAACTCACTTTTCACCCCTTAAGCTTTTAAAAAC

...

234096 AGCAAAAGCTCATCTCTTGAAGAAAGAGTACCAACAAGTTTAAAAATAAGCGCCGCCG
|-----|
234146

234156 CAACAGCTTCAATATTTTACTATCATTTTAAGCTTACTTTTTTAGC

...

234329 TGCTCTTTTGGCGATTTTTCGCACTTTGAAAAATGCCCTAAAAAAGAGTAAAGAGCGCATTT
|-----|
234379

234389 ATACTATGAGTTTCATCGCAGGTTAATAGATAATTTTATACTTTTGCARAGGCATTTAAAAA

...
222848 TATATAATACCAAGATTTTGGACCAAAATTAAACTTACCTGCAGGTACAGGAGCTTTGAT |-----|
222898
222908 TTTTACTTAAATAACCACTACTACTATCATCATCCATCTGTATATCATAA
...
222967 TTCTTTTAAATCCAGCATAAAAACATAGGGATTATAGAGATGATAAACAATAATCGTTGCCAA |-----|
223017
223027 AAGCGTGTAATTTTACAAACCCCTTACAAAAACTCCACCCCCAAA
...
223075 CTTCTCAACATCATCACAAAACAGTTGGCACTAAATCACCACCAAGAAATCGCTGCCGA |-----|
223125
223135 TAAAAATGAAAAAGTCTTTCAAATAATGTAAAGCATAAATTTTCTT
...
223203 GGTCTATAACCATATAATGCAAAAGCAATGCAGACAATACCGCAATAACAGCGATTAAA |-----|
223253
223263 GCATAGTAGCTATATATTCACCTTGCAAAAGTTGTCCAAAGTTTGC
...
223369 GTCCCATTTCTCGCTACCATAAGCAATAAACTACTAGCGTTAATAAAAGCGTTACAA |-----|
223419
223429 ATACTTTTCTCATAGCGTCCCTTTCTCTGTGTAAGCAATCTTCCATTCCTCAAACTTTCCGA |-----|
223443
223489 AGGATTTTGA
...
223799 TCATGCAAAACAATACCATCAGCAGCGATGTAAAAACAATCCACAGGACAAACGATAGAA |-----|
223849
223859 CAAGGAGCATCATCACAATGCATACAAAGAANAATAGAAAGTTGAACCTT
...
223963 AAGCTCATGTGCTTCATCACAAGCTACCGGCAACCAATTACATCAATACAAAGCTCATTT |-----|
224013
224023 ATCACAAGAAAGATTTCAAGCGTTCTTTTCTTAAATTTGCAAAATTTACCTTACTCATTTT |-----|
224040
224083 CCATCCTTACTGA
...
224504 AATAATCCTTAGCTTTTGCACAGAAACAATAATTAGTATCCAGCGGTAAAGCGCTTTT |-----|
224554
224564 GATCTTCAAAACTTGGATATTTTTCACCAAAATCGTTTCTTGGAGAATGTAAAGGCTCAC |-----|

224595
224624 GGTGGAGTGAATTTTATCTTTAAAAG
...
225461 GTTGTGCAGTCGGAAGTAATACCATTTTCCCATAACTACTAAAGCTTTTAAAGGGTTCCTCG |-----|
225511
225521 CATTATCGATAGATCATTTTCCATCTTTTCCATTTTAAAGCCGCTGC
...
225626 CTAAAAGCCACTCATATAATCCACACCCCAAAATTTTGGCATAGTATCTTCCAAAGTCGCTTCAT |-----|
225676
225686 TTAAAGGATAATAACCTGGTAAATTTTCGCTCAACAAGCCATATC
...
226108 CTTGCATAAATATCCGCTTTTGTGCACTTTTGTAAATCTAGGATCTACAAGCATAAAG |-----|
226158
226168 TTTGCACCTTTTCTTTTGTCTTTTAAAGAAATGTCTAAAACCCACAG
...
226240 ATCATAAATCATATCTTTACTTCTTTGGATGTCTCCAAGATGGTTTGTGCATAGCGCCATAA |-----|
226290
226300 CCAAAATGATTCGCCACACCGCGACTGTTGGCTGTGTCAAAATCTAGCTTGATGATCT |-----|
226320
226360 ACATGTTGCTGTTCCAA
...
226588 ATATCAGATCCCTTGCAGCAATGTCACCTGAACTCACAGGATGATCTTGAGCGATTTCT |-----|
226638
226648 TGACGCAACCAACACACCATTTTTCACCTTCTGCAATAATCCACAGC
...
226791 AGCAAGCAACGGTGTGTGCTAGGCTTGAAAGTGTGCCATTTTAAAGAAAAGACGACGCGT |-----|
226841
226851 AAGCTTGATATTTTCACTACACTTGACATTTACTCAGCCTTTTATAT
...
227074 GTCCCTCTTTGTTTATTTTGTCTTCTTCTTAAATAATCTGAACGGCATAGATCGTTTTCG |-----|
227124
227134 ACTTTTTTGAATTTTATTTTTCAGGTTTTCCTCACTTCAAGCAAA
...
227206 GCAATACTTGCAATTTTCTTTATATAATAATTCGATTTTATTCATGAATAAATCGCTTAAA |-----|
227256
227266 AACTCATCAATCGCTTTTATTGATCACTTTTTCAAAAAGCTCTTTGG

212060 CATAAGAACTATACTCAAAACACTTGAGTTTTTATCAGTAGAAATCAAAACGCTATTA
212110
212120 CAGCTACCAATAAAACACCTAATTTACCTAAATTTAATACAAATG
...
212561 CTCCTTCGCCCATAGATAAAGCTTGCGGTTTATTTCTTACTATATTCATCGCTTCAC
212611
212621 CAAACCACCTAAATGATAAAGCATACCAAAAGGTACAATAATAA
...
212702 TCCAAACAACCGCCTTATATCCACCTAAAAAGTATAAGCAACAATAATAATCGTTCAG
212752
212762 TAGTTAAAGCATATTCGTATCTGATACCACCAAGTACTTTCAAAAG
...
213266 TCAATCTCTCCTTGATCATTAAGATATTGCTCTAATGCTAAATTTCCATACGATGAA
213316
213326 AGATATACTTAAAGCTTTTTCATTAAATAATTTAAAAAGTTCAAAACGACCATTTAATAA
213370
213386 TGGCTTTTCTCTATATATGATTTTAGCAAGCTTTGCCAGCC
...
214603 TTAACATGGCATAAGCAGTATCCTCTCCTCTGTAAGAATTGAAAAATTTAATCGCTTCAT
214653
214663 CAATAAGTAAATATTTAGAAATATCATTCTCTTTTGAAGGTAAAGAA
...
215295 GTTTTAAATTTCTTAGCCACACCTATTTTGTGCTAAATAAAATCCGTATCGCTTTCA
215345
215355 TTGCTAAATTTTACCTTTCATAAGAACTTATGATATTTGATTCCT
...
215507 AAGATTGAAAAATAACGCATAAAATATCTTCACTTGTATTTTCATCAAGCGCTTTAC
215557
215567 AAGATAAGCAANTGCAATTTGTTAAATGTGCTTCATCGCAAACTGGG
...
216572 AGTATATCTATAAGCTCAAAACTTGCATTTTTTATCTTTTACATCTAAAAAGAACGATCTAA
216622
216632 AAGCTCTATAAGCATAACTTTTATTTTGGGATTTAAGAAAGTTT
...
217612 AAACCTACTATATTTAAGGCATTAATTACCTCTGCACCCATAGCAAAAACATCGCTTAA
217662
217672 GCATTTGCTGCAGCAANTAGCACCACCAAAATGATAAGCACTATCTACTATAGGAGTATRAAA

217732 TCAAGCGTTTGCACCAAAAGCTAAATCGGGCTGATTTTGATATACGCTTGCATCTTCATT
217735
...
218398 ATAGAATTTAAACAAATTTCTAAAAATTTCAATTTCTTGAAGTTTTCAGAGCGTTTTT
218448
218458 GTTTCAACTATAGGCTGAGGACAGATAGATTGCCACAATCAATTT
...
218697 TTTGCAACTTCGACATTTTCCCTTAGTTACAGTTTCAAGTTGTGCTATAGCATCGTTAATC
218747
218757 TGTGTGATTCCTGCTGCTGTTGTTCTTTTACAGATTTCTGCTACTTCAT
...
218922 GCACGTGCAGCTTCAATAGCTGCATTTAAAGCAAGAAGATTCGTTTGTCTCAGCGATCT
218972
218982 TTAATCACATGACAATATTTTTTATATCATCATCAGCTTGTGAAGCAA
...
219922 TAGGATTCATATCTTTTAAATTTTAAACTAGATTTTATTTGGATGTATCGCCATAA
219972
219982 AACCATCTGAATTTAAAGAACTCTTAANTCACCATCATATTTTGG
...
220473 GAGGTTGATCTAAACATTTCTGTTAAAGTACCTGAAATACTTTTATTTAAAGCGATCAT
220523
220533 TCCCCAAAAATTCCTTGAGTTTGAGTAGCATATTTCTTTCGTTAG
...
220812 TATTTTGTAGAGAATTTATTTTTCCTTTTAAACAAATCTTTAAAGCGTTTATT
220862
220872 AGCGTATTTCTTCTATGCTTTTTTGCAAAATAGAGTAATTTTCCATCAATTCATAGCTCT
220932 AGGAGTAAGTTTGTTCCTGATCTTTTAGAGCGCTCTGCTCTGTGCTAACTAAAATTC
220961
220992 TTGAGAAATTTTGTGAAGGTTTGT
...
221574 TAAATTCCTCCACTATAAATATTGATCTATTTTCTCTAGCAAAACCAAGCGTTAAAT
221624
221634 CAAATTTTCTTGCTATCATACCCCTAAACAAGTAGGAGCGGTGCG
...
222089 GCTTTAATATGCACACTCATACCATCATCTTTTGTGTTCTATGCTCTCAATATCGTTACT
222139
222149 TTCGAAATGATATTTTTCACTCATTAATAATATCCTATTTGCTAATGCT

```
...
203634 TCCTCACCCACCTTTATCTCTTACTCTCTTTTATATAAAGTTCTAACAGCAGTATCGCTCACA |-----|
203684
203694 TGTTCACCTATCCAAACATTTTCTTAATATCTTTCATGCAAAACCA
...
204122 AATATTTTAAAGCCCTCATCTCCATTTTGAGCTCCGATTACTTTACTAAAGCGTTGCGCT |-----|
204172
204182 TAATATATTAATCATTTGATTCTCTAGCCTTAACCTCATCTTCAACT
...
204851 AACATGGCAATACCGAGCTCATTTAACGGTTGTCTCCATTGATGTCTATATATCGCTAATC |-----|
204901
204911 ATTTGTTCTAATGAAGATTTCAAAATCTCTTCATATGCTATTTAA
...
205477 TGAAGCCCTTAAACCTATATAGGCAAGATTTATACCGTAACAAACAATAACGCCATAC |-----|
205527
205537 GATAAATAAATAAATAAATTTTCAAAAAATCACTAAAAAACTGTG
...
206050 TTAAGCCACCAACACATAGATAGCAGTGTCCAAAACCTAGATAAAAAAGCAACGCTGATA |-----|
206100
206110 ATAGCTAAAAAATCTATACTCACATCAATCAAAATTCATAAATTCCTTTAA
...
206199 TGTTGCACATGAATAATAGGATAGTCTTTATATCTTACTCTTCAACATTTATCGCCAAAT |-----|
206249
206259 AAATTTCTATGAGTGATATGAGCCACTTGAGCTAAATTCCTTGGA
...
206676 ACAGCAACTTTTTTCTCAGTTCTTTTAGCATCATTTAAATTTTGAAAACTAGCGTTCCAA |-----|
206726
206735
206736 ACGCTTCTTATAGAGATAATTTTCTTAGTGCTAACTTCTTTACAAAAATTTACTTCA |-----|
206740
206796 TCAATTTTGTGATTTTCTAAGCGCATTTTAAATCTTCTTTATTTGAAATTTCAGTT |-----|
206819
206856 GAAATFACAGCTCTTAAAT
...
208378 ATAAAACTAACATATAAATTTTAAATCATCTTTTCTCCTTTGAAAAAGCAGCTGAT |-----|
208428
208438 TGCAATTTTCATAAAAAATTTTCTTCAGCCTTGGCAAGTTCTTTT |-----|
```

```
...
209121 ACCATTAAAGCCTTGCTGTATATCTTACAGCAACCGCATCAAAACCATCCACAGCGCTCTTTT |-----|
209171
209181 GCGTCCGTTGCTTACGCCAATCTCTTTTCCAAATTTGGGCAATTTT
...
209463 CCAAGCATTCTTCTAGTATCTGTAATATATAGGAGTTAAAAATTTCAATTAACGCTTTAA |-----|
209513
209523 ATCAGCAAGCAATTTCTTCAGCACCTTGAATTTCAATTTCAAGCATT
...
209734 TCTTTTGCATGTCTATTAAAGAATGATGTTTTAAATTTAAATGAGCTCTATCGCTGATA |-----|
209784
209794 TATAATCTACCTTTTAAATTTTCAAAATTTGAGCCATTTTCAGCAATCA
...
209986 ACAAATCATATAATTTTCACATAATTTATCAACTACCTTACCTTTTCCTTCATCGCCCAT |-----|
210036
210046 TGAATGCCGACATAATATCTGCTTTGCTCATAAAAATTTCCCTTGA
...
210126 GAATCCCCAAAGAATAGTCAACTATCTTTTCATCTGATTTGAAGTCAGTGGCGATCACC |-----|
210176
210186 AGATACGCTAATCTTCCAAAGTGATAATTTATAGCGATTTTAAATGAAAAACAAATAAAT |-----|
210217
210246 TTATCAATAATTTCTTTTGACTTCATAT
...
210394 CTATGATTGCACCTTGGAAATTTTAAGTCAGTGATTTTGGGATTTTGGTGGCTTTGGG |-----|
210444
210454 CTGTTACTGAGAAATGACACGTTGGGTTGGGAGTTTTTAGAGCT
...
211114 TTTTATTTGCTAGAGATAATGCTATAAAAGCGCATTAATCGGTATGATGATCGCTCTT |-----|
211164
211174 TAATCGCTTTTGGCTTTTATCTTGCAGGTGCATCTAGCAAACTCATAGAACTTTCTCC |-----|
211176
...
211928 CAAAAATCTTCCAAAAGCACACAGTAGCAGCTCTGTTATCATACCTTAAGATCGCTCCTA |-----|
211978
211988 CTCTTGTCTACTCTCCAAAATAAAGAAAAAGCATTTACAGACC
...
|-----|
```


...
193421 AAGCAAGGTTGGTTTTTACCTCTCAAAATGTTGCTACTTCAAGTTAAAGACGACCAAA |-----|
193471 193480
193481 GCGTTTGGCTTCTGGGTCTAAACAATCCATTCTCGTTTACTTCGTTGGCTTT
193485
...
193667 CAATAAACTATAAAGCAGCACCTTATCTTTTTCGCAAAATTCACATAATGGCGATCTTA |-----|
193717
193727 TAAGTTCCTACCTGCTCTTTATGTTTATCTTGCATAGCAAGACCTG
...
194105 TGAAGGTAATTTTTTCAAAATTTTGCAAAGTTTGAATGCTTCTTTTTTAACGCTTAAA |-----|
194155
194165 ATTTGTTCTTTGCTTCTFAAAAAACCTCTCTCTTGAGAAATGTATAT
...
194600 AGCTTTAAAAAATCTCCTTAGTTTGTAGCATATTCGATAAATTCATTTAAAGCGATATCA |-----|
194650
194660 TTTAAATCAAGTTTAAATCTTTTATTAAAAAAGTACATTTTCCA
...
194864 GTAATGCAGGATATTTTGTCTCTTAAAAATCCAAACCATATCTGCATCGCTATTT |-----|
194914
194924 TTCCAACACAAATACAAATATCATCATAGAATAATTTTAGATC
...
195118 GATTTTTCTTTAAAGTTTCATTAACAAAATCAACCAAAGCTCTTTAGAACGATAATT |-----|
195168
195178 AGTATTTAAATTAATCACTTTTAAATTTGACTAAATCTTGTTTTAAACAAATCAATAATTC
|-----|
195238 TTTTTTCCTTTACGAAGGGATAAATACTTTGTTTTTTATCTCTCATATAAAAAAAGT
195255
195298 GCGATTTTCTTAACCCCTTCTCTGAAACAAGCTCAGCGATCAAGGGCGTAAAAATTTG |-----|
195334
195358 ATATTGGATCAGGCTTGATCTATCTTGAAATTCAT
...
195802 CTCATAATAGCTTTATTGATATAGCTTTAGTTGGATTAGAAATTTTAGGACGATTTTG |-----|
195852
195862 AAAATAAGCATTCTCATAAAAATTTTCAAGCTCGTTAAAAAAATTT
...
196971 CGCTTTCCTTCTCGCTTAAACCCCTAGGTTCTGCTAAATTTAAAAATATAACGATAATA |-----|

197031 ATAGTTCTCTTTTGTATTTTCCAATAAATTAACGACTTAAAGAAAGCGGAAACTCGAA |-----|
197063
197091 AGGATTATGCTTTTAAATAGGAGCTTTT
...
198807 GCATAAAACTTTAGCTTAAGCTCCCATCTACCTTTTATCTAATTTAGGTAAGCGATACTT |-----|
198857
198867 AAAATTCATTTTAACTTTGCTTCTTAACCTTTTGATCAAAATTTT
...
199147 AGTCAAAATAAAAAATCATCTCATACAGGATGCTGCTATATAAAATCGTTACAA |-----|
199197
199207 TACAAGCAATAATTATAGCCAAAAGTGAAGCAAAATACCATAGG
...
199636 TCTTTTGTATTATTAATAAAGTTTCAATTTTCAGTTTTATCACCACCAATACGCCGATA |-----|
199686
199696 AAAATATTACTTTTTTGGTAAAGCTCTTTTCCCTATAAGATTAACT
...
201283 CAAAAACTATTGATAAATGTTCCATCTTACTTTAAGCTATTAAATAAAGCGATCAAA |-----|
201333
201343 GCAACTATTTCTTTGATTTCACCTTTAGCAAAAGCATCTTTTACTT
...
201693 AAGAAAAGGCTATCATATAAGCAAAATTCACCACCTTACAGAATACATACCATCTGT |-----|
201743
201753 TTCTGATTTAAATGGACGAATAAGCTGTGAATGACAAGCATTTACAA
...
201958 TTTTCTAACCAACTAAACATATTTTCTCCTTTTATGCTGCCATAGCGAAGCGCTTTT |-----|
202008
202018 GGTTCTTTTATCAGCACTCTTCCACAAGCAATTTGTTGTAATATTATAAGTAAACATA
|-----|
202078 AAGAAACCGATAAGATACACAATCCCAATCTCTTAATCCAATAATAAGGAACAATA
202108
202138 GCCACAACAGTATCAATAAAGCTATAAAGCAAGTTACCATCTCATCTGCTGCTCCAC |-----|
202186
202198 ATCATACCTTGAGTAATACCCGCAATCCACATTTGAAGCAAAATA
...
202508 GACTCTCGGAAGCTGACTCCACTCCACTTTTCATAGTAGCAAGATATTATTCCTGAAC |-----|
202558
202568 CCCAAGAGGTAAATCAATACTACAGAGAAAACCGAACCCATAGT

185217 ATCATCAAGTTCTAAAGACTAATATATTTTCAATTTAAACAAGCACTTAAACGACTTTT
185267
185277 AAATGCTCTTTTTTAAGTATATTTAACTTTTTATGAGAACTTAA
...
185651 TTAAGACTTGGAGCATTACGCTATCTACTATCTTGGAGTGATAATAAAACGATTTCG
185701
185711 GTTCCTTTAGAAATTTAAACTTCTCCTTTAAACAAGAACCTAGGA
...
185863 AGAAGCTTTTTTGTATGGTATCAGGCAATAGTTCTTGGTCTTTTTTGCTGCTTATT
185913
185923 ATCTTCTGGATACTTAAATACACTTAAGCTTGGATTAAATCCTAAGC
...
186059 GTCCCATTTTCAGTACCTTTAGAACTTTCTTAACTTTGATAAATTTATAGTATCGCCTACA
186109
186119 GAAATAATGGCTTGTGGTTATTAAAGCCATAAGCTTTGGATTAG
...
186355 ATTAAACTCACAGCAAGTATGCTAACATCGATAAATACCTGTTTTTAAGACGATTTTC
186405
186415 TAGTTTTTGCATAAATTTTTTCACACTTTTTTAAGTGAATTTGTT
...
186602 TTGCTCTCGATCATCAAAACTACCGCTATATTCACCTTTGGCTTGGTTAGCGCTACACA
186652
186662 GACGCTTTTAGTTAGCTTTTGCTTCTCTGTAGAGAGTGATATAAATCTTTAAA
186662
...
186916 ATCAAGACTTTCTTGTATGCTTAATTTTTTCGTTAATACTGATATCAAAAAGACGCTTTTG
186966
186976 ACAATCAAGCATAAAGATAATATGGCAAAAAGAAATGTTAATTAAT
...
187360 TTCGACCTCATTTGCTATAATACGCCATAGCAATACGCTTAAGCTGTCTATAGGACGCTTTG
187410
187413
187420 AGCGTCTTTTTTATTTGCGTTTCAAAAAGCTCAGCAGCGTGTTCAGGATTTGCTTTTT
187420
...
189083 AGTCCATCAGACCCCATCAATCAATAGCGCTTGTAATTCATTTCTACAATAGCATCGCCTTTT
189133

189143 TTACTATAAGATTTTATACGCAAGCTCTTTCATGTATTTTGTGCGT
...
189404 CGAATTCCTGCTAAACATCATATAAATTTCCAAATAAGCCGTACCGCAACAAGCGATAAAA
189454
189464 TCGCGAGTTGAACAAGATAAGTTGAACGGATAGTTTTTTTAGAAA
...
189946 TTTTTTCACTGATTTTGGCATACATCAAGAAAATATTTTAAACTAAATGGCGGATAAAG
189996
190006 ATAAACTTTTAAATCCCTTACCCTTTTCCACCTTGGCATTTAAGATAG
...
190162 TGAATTTTCTTGCATATATTTCAATTAAACAACATCAGGTAAGCATCGTAAACGATTTGCT
190212
190222 TACTTCTCTAGTTTGAAAATAAATATCATCATTTTGTGCTGTTCCA
...
190935 ATAATTTTATAAATTTTTTAACTATCAATTTACTATAAATTTCTTAAAAGCGCTTTTAAC
190985
190995 GAAACTTGAAATTAGTTTAATCTAGCTATATAATTAACAGCCTCA
...
191107 GCCAAAGGAGTGATATTAGCAGAAATTTGCAGCTTGAATATCTAATATTGTATCGCCTATC
191157
191167 ATATAGCATTTTCTTGAGTTTATTGAGTCTTTTAAAGGCTAAAA
...
192390 TTAACATCAGATAATGCTCTTAACCCAAATTTCTAATTGATCAAGCCAAAATCGTCAAAA
192440
192450 TGATAACCAAGTCTAATCCCTGGTGCATAAAGATTTATCCATATCTAAATTTACCTTCAAAG
192478
192510 TAATTATAGTTTAAAGTTGGAGTG
...
192684 GAGTTTTTAAAAAATTTGTAGAAAGTTGGAAAAATCCCAACCAAAAATATTAAAGTTTAGA
192734
192744 GAATTGTGGCTTCTTCTGCTTTGGCGGCTCCATATTTTTTACGCTCAACAGTTCTACT
192770
192804 ATCTCTAGTAAGAAGTCTTTTA
...
192944 TTGAAGTTTCTTGTTTAGTCACAAGTAAAGGTTGAACCACTTTTAAGCTTTAATCGCCTCGT
192994
193004 GTCCACCAAGCCAAGTGTTTTAAATCAACACCAATTAACACTAATTTT

```
...
177947 TTAACCGAAGAGATATAGAAATATCAAAAGTGGCTAAACATGAGAGTGTAACGCCAAGT |----|
178007 GTAGAAGAAAAAACAACCTTATCGGGAATTTATTACAGCTAANA 177997
...
178582 TCAACTAGAAATATACTAAATATATGGCTTCCACTTTACAAGAAATTTGCGCAGCGTTTCC |----|
178642 CGACTTCAAAATACTGGAAATTTTCCAAGATTTAGAAAAATTATAAC 178632
...
178722 AAAATGCCACAAAAATAGCCCTTGTAAAGCTTCTCAGACCTTAACAGAAAGCGTTAATA |----|
178782 ACACTTTTGCAACCTTAGATAAAATCCAAAAAAGTAAATGATGA 178772
...
178969 TTCAAAGCTAGTTAGTGCAGTTGCAAGCAAAAAATGAAATCAATCAAGACAATCGCCTGA |----|
179029 TACRACCATCACTGATCCAGGACATCAATACAATCTAAGTATAGAA 179019
...
179427 ATAGAACCATACAAACGAGGAAGTTTGTATATAGTCATTTATGATGATAAAGCGGATAAA |----|
179487 AACTCACCAAAACCATCACCATAGATGTAATAACAACCATGAATGA 179477
...
179535 TCATGCGCCAATCAATGCTAATACAGATGATGAATGACAATAAATAATTCAAACGATGATG |----|
179595 TTGATGATCATATCAATGCAAGTTTTAGCTACGATGCAAAAACTGG 179585
...
179928 ATTACCGCAAACTTACAGGTAATAATAGCCTCAGATGGTGAATAAACAATGTCGTCAATT |----|
179988 CAAGCAATGAACCTTTTATAATTCTGTTTATAGCGAGTACCAATC 179978
...
180060 AATTAGCGCCCTTATACAATATCAATCAAGTTACGGTGCAGCAGCAAAAAATCGTCACAA |----|
180120 CCGTAGATCAAAATGCTTGTATACCTTACTAGGGCTTAAATCTTAAGC 180110
...
180614 TGATTTTTCTTGGAAAAATTTGGCAAAACACCCCTACTTCACCTCTTTAAACGCTATAA |----|
180674 TATGTAATCAAGATGGATGGTTAGAAAAAGATGAACCTTGATCTTGGA 180664
```

```
...
180851 TTTAAAAAAATTTAGATGCAAAATGATCCATAAAATATGACTTTGCACCTCTATCGTTTAGG |----|
180911 GCAAAAGTAAGGAAATTGACAAATTTAAGGAGTGAAGAATGGGTTTT 180901
...
181066 TAGCCTGTGGAATTCCTGGCATTTATCCTTTAGCAACAACAACTCCAAGCGTCTTTG |----|
181126 GTTCTTTTACGGCTACTTTTAACTTATTTTAAATCCACCACCTCTACC 181116
...
181612 CAAAACTTGTATTAANAACATAATGGAAGTTTTATAAAACTTTATTTTAAATCGTTTAG |----|
181672 GAGCTACTATTATTAAAGTAGCTTGGGATTATTTTAGTTAAAAATTG 181662
...
181728 TAAGCTCCACATAGGTAAAAATATACCTAAAGCCAAAAACAAACTAAAGCGCTACAA |----|
181788 ATAAAGTCATCATAGGCTCTAAAAAAATCAAAAAATCTATCCATCAA 181778
...
182603 CTAATATTTTTTTCATAATCTTGCAGAATGCATAAAATATAATTCCTTAATCGCTTCTT |----|
182663 GAACACTTTAATCCAACCTTCACACACAACAAAGCTAGCTCTTTAAAAAA 182653
...
182944 CTTGTTCAAACATAGTTTGAAGGCCATATTTTTTGCAAGCTCCATTTAAATCGTTTAT |----|
183004 CTTGATTTTCCCTAAATTCACACAAGCTAAATCTTCATTGATCTCTAAGCACTCAGCGATTA |----|
183064 ATTCTCTGCCAAAAAATCCACTATGATGACAAGCTCACATCTCTTTTGGTTCAAAAAATT |----|
183124 CACCTTCAAAATTCAAATATAATGTTTTTTTGGCTTTTACACCAAGGACAAAGTTTTC 183094
183184 TAACCAAAACGCTGTGCTATAACAAGACTCAAAAGCGATGCTATCAAGTAACCTTTTGGCTT |----|
183244 GC 183190
...
183548 TGTGCTTTTACCACCTACCTGTAGGACCGGTTAATAAAATCATTCATTGGGACGATGTAA |----|
183608 AATCTTTTTTAAAAATTTCTAAATTTTATTTCCAAAGATTTAGTTTA 183598
... |----|
```

172134 TAACTCTTCTAAATATCCCAATATAAAGCGTCTTTGAACCTTGCATTTTAGCCTTAAACCTC
172160
172194 CAAAGTATCTTTAAAAATTTCA
...
172231 GCCAAACCTTCTCCTGTTTCATANGCCTTAGCACCAGATTAAAGCCCATGGCGCTTATA
172281 |-----|
172291 TCTCTTGCTTTAAACGCCCAAGCTCAGCACTTAAATCCTCGCATAAAAACACTTTCCA
172304 |-----|
172351 AAGGATAA
...
172461 TTGCTAAATCATTTGCAAAATGTGATACTCCACTCTTTTACTCAATAGCGATTAAA
172511 |-----|
172521 AAGCTCACTTTGAAAGCTGAATTA AAAACTCTTTCATTTTACA
...
172581 GCCAGCAAAATTTCAAGCCCTTCTTGTAAATATCCCAAAATTTTCCAAAACGCTGATA
172631 |-----|
172641 GCCAAAATAATTCGCAAAACCTTGTTATCCAAATTCATCAAAAGCTTGTCTAGTTTTAA
|-----|
172701 AGCATCGCTTGGTAAAACTTTTTTTAAGCGTATGAAAAAGAAATTTCCCTTTTAAATGGC
172704 |-----|
...
172999 TCACCCCTACGGCTAAATTCATACAAAAGGCCTTTCACGCACACAAAATATCGCTATTT
173049 |-----|
173059 TTACTAAAATACGCATTAAATAGGACTGTGTTTTTAACTATACAAAG
...
173498 GTATAAGCAAGTAAATGCCCTTTTTTAAACCTTTTCTAAAAACAGCTTTTATCGTTAATC
173548 |-----|
173558 TTTGAAATGATAGTAAGACTAAATGCTCTATTTTATCATTTGCTTATAG
...
174307 TCATCGATAAAATGATTTTTTTAGACAAATACAAATATCTTGAAAAATTTGGCGATGAAG
174357 |-----|
174367 AGGATAAAAAGCTTTATGCAAAATATGATTATTATATATTATGCAACTGAAATTTTAGACT
|-----|
174427 ATCTTCCTGAATATATGCGGATGGAGTGTTTTTTATTTGGCTTCAAAAATGATAGAAAAAT
174444 |-----|
174487 ATCCTGAAAAAAA
...
174531 GAAAGGGTATTTTTTATCATCTTTGAAGCAAAAATGAAATTTTAGAAAAACAGATCAAA
174581 |-----|

174591 AACCTACAAAATGAATTTTTAAATTTAGGGCTTTTTTGATAAAAAATA
...
174829 TTAATCTAAATCCTACGCTAAGCTATAAAATAGCCGAGATATTGTTTTTGACGCTTATG
174879 |-----|
174889 ATACGGGGCTGATTTTTATGCTAGTAAAGAAAGAAAGATTTTTTA
...
175137 TTTTTTAAAGATCATTTTGTGAATGAAAAAAATTTGTGATTTTCATCATTTAGCGATGAAA
175187 |-----|
175197 AAATTCATACGGCAAAACCCCTTATTTATCATCTCTAAAAACTCTCC
...
175317 CAGCTTTGCAACTTGCACCTTGACAAAAAGACTGCTTGAAGAAAAAATTTGACGCTATTG
175367 |-----|
175377 CTTTGGAAATTTAAAAAAGATTATGAAAAATAAAATCGATCTTGGCCAT
...
175429 TCTTAAAAACAACCTCGTAAAGCACTTAATGATGAATGATGAATTTATAAGCGTTAAAG
175479 |-----|
175489 ACTTTTTTAATCACAATGAAAAAGAAAAAATTTCTTCAAAAACA
...
176164 GCTTGATCTATCATCAACAATAGATTAAAGGAAGAATCCATGAGAGTTTAAACGATATTT
176214 |-----|
176224 TTACTCTTTATGACAAGCATTTTTTGCAGTGCAAAATCAAGGATGTAG
...
176288 GAGATAACCAACTTATAGGTTATGGACTTGTGTAGGGCTTAATGGAGCGCGATGAA
176338 |-----|
176348 CAAGCTCTAAATTTACACTACAATCTATCTCAAAACCTTTTTCGAAG
...
177238 GAGATCATCTAATGAAAGTGGATAAATTTTTTAAATACTTATAGTATGAATAGCGCTTTAC
177288 |-----|
177298 TTGATCTGCGCCCAAGCAAGAAAGTTTAGAATCAAGCATCAAAAT
...
177433 TTTTAGATGTATCTTTTAAAAAGTCAAAATTCACCTTTTGGAAAAAGATGCCAGCGATGAAA
177483 |-----|
177493 TATACTCTCCATGTATAATGATACCATGAGTAAAGCTTTAAGTGG
...
177681 TGGCAAAATACCGCATTTAAATACAAATAGAAATAGATAAAGAACTAAACAACGATACTC
177731 |-----|
177741 AAAAAACGGAAATGATAAAGCGAGTAAAATTCGACAGCAGATTAA

165738 ATACTTGAATTTTCACTGGCTAATCTTATATGCTTAGTATCTTCATGCGAAGCGGATAATA 165788
165798 AAAGCTAAATCTCTCTCATCAAAAGCATATAAAAGCACTAGCTGTAT
...
166083 TCAAAGCGTGAGTTATGCAAAATAATGAAGAGGTAAAACTCTTAATCTTAAGCGTCAGTA 166133
166143 TGGCTTTAGCTTTACTTTTTGAAGGCTCTAAGCTCACACAAAATGAT
...
166595 TGACAAATATCGAAATTTTAGAAAATATGCTCAAACTCCACAAAAAACTCAACGATGAAA 166645
166655 CAACGGCTTAAATTGGGAAATGCTACACCAAGAAGGAAAACTGATCAGTTGGAGAC 166712
--|
166715 GCTGTATTATATGGAATGTCAGAACTTATAGACTCTTTTACATGGAAACAT 166717
...
167385 TAGGAGTACTTGGCTTTTCGGCTTGCTTGATTATTTTAATCTTTATTATAGCGTTGGAG 167435
167445 AAGATTCGTTTGGCGTGGTTTATGCTCTTCATTTTGTACAAAT
...
167538 TGGCTTCGGTTTTTTTAGCCCTTTTTATCACTCTCTTTCTTACTCCTTTTATCGCTAAAG 167588
167598 AAATCAACCAAAAATATTACCATTATGATAATACAAAGGAAGTTTC
...
168470 AATTTCATCTTTACTTATAAAGAGTAAGGGCGTAAAGTGATATAAATTGAGCGTTTTC 168520
168530 ACACCTTTGCATACCTTAAATTTTCAGCCAAACCTGGCGCTCGTGAAGTTGCATCAAAAA
168590 CCATTCAAATTTGATCGTTTAAATGATGGCTAAGATAAGATGGTCATAAACCTTGTCTCAAG 168602
168650 ACAAAATTT
...
169095 AGCTCGGATGAAGAACTCTCCCTCGCTACTTTTTTCCTGTGCAAACTACTAAACGCTTTTG 169145
169155 CCTACTTCTTTTCAAAACTTTTTTCTATGACTTCATCAACAATT
...
169243 ACTTGCTCCATATCAAAAGCAGCGGTATCTTCTTCTTTAGAAATACGCAAAAACGCTTATC 169293
169303 CTATCAAAACCAAAATCTTTTAACAACTCACAAAGCTCTTCAAAAT

...
169777 CCTTTAAAACTTGGGATGGCAGCAAAAGAGCATTTTTTGATTACAACCTTCGCGGATTTTT 169827
169837 ATAAAGCATGGGAATTTGAACCTGTGATATAATACGTTTTAGAAATTTT
...
169962 ACCTGTGAAAGATCAACCTCAGGAAGTCTTTTCATAAGTCTTCGCGGTACGCTGCAT 170012
170022 CAAACAACCCGTAACCAACCAAAAGAGAATCTTTTTTTCTTTGTTTCG
...
170546 TTAATGCTCTAAATGACGCAAGCGAACTTTTCGATCTGGCCAATAGCGAAAACGATACAG 170596
170606 AACACACTTGAAGCTTTTATTTAATGACGACCAAAAACCTTGAAGATAC
...
170871 AGTTAAAGGAGAAAATGCTTATGTTATTAAAAGCTGAAAATGCTATCATCTGTTTGGT 170921
170931 AAGAACTTCTCCTTTTGATAGTCAGAGACGCCGTGCATACAAGCTTTTCAAGTGTATGGT 170957
170991 AAGCCCTGAACCTTGATGATGAT
...
171347 TTTCTCAAGTGGATAAATATCCTTGATGGGATATCAAAAAAATGATAGAAGCGGTTTGA 171397
171407 TTGCTTTAAAAGCGGAGTAACAGGCTTAAAAAATAGTAATTTCTGC
...
171555 TTTCTTAAATCTTTATCGCAATAAACACTCACAACTTTATAGCTTTTATCGTTTGA 171605
171615 TTTAATAAGAAAAACATAGTTTGAGTATGGAGTTTTTATCTATTTCTA
...
171747 TTGGATATTTTTTCATAAATCTCTTGATTACCTAAATCAAAACATAGCTTGCTGTGCCA 171797
...
171807 TTGTTTTTTAACTCGACTAAATCGCCATCGCCACAATCTACAATCAAAAGTCATAGCTCCT 171827
171867 AAGCCAAAGCACCTGTTAAAA 171833
...
172074 GCCTTTTTCAGAAAAAACTCTATACAAAAATGAGCTTTTCTTCATCATACACGCCATTT 172124
...
|-----|

```
...
157989 TCATTAATAATCCCACTCTCATTTGCTAAGTAAGTATTAGNAAAAAGCTTAACGCTTTCT
      |-----|
      158039
158049 GCCTCATCCGAATCCATATATAAAAAATATCAACATTTTTTTTCTCTA
...
158656 ATTAATAATCCTAAAAAATCCCTAGGATTAAAAACAATTTCTTGCCAAAGCGATTGGA
      |-----|
      158706
158716 TTTAGTATGTGAATCATAGCACCAGATACTCTAAACCTAAATTT
...
159425 TTTAAATGCTTATAAAAAGCTGATGTTCTTATCTTTATTGTCATATAAGCGTTTGA
      |-----|
      159475
159485 TTGGGTTCTATATGATAAAAAATGATATATATAAATTCCATAAAAAAC
...
160115 TCGTGATCTTTAGATGAATATCCTTCAACGAAAAATAAATTCTATATTCATCGCTTTGA
      |-----|
      160165
160175 AAAAAATTTAGCTTTTGAAGAATCTTTCTTCTATATAAGGAGCTTCTTTACTTAATCCA
      |-----|
      160215
160235 AAAGGAATAATAACAGATAAAAGAGGCATTAAACAT
...
160396 TTAGGATTTATACTCATGATTTCTTTGGCCTTAAATCAAAACCTTGGTTTATCGCTTGCT
      |-----|
      160446
160456 TTTAAAGCAGCAGCAAAATCCCGCTCAGTGATAATAACCCACTAATT
...
160510 TCAAGCACCACACAAAGTCCTAGTTTTCCACTAGTCATCACATCAACAAGATCGTTAAAT
      |-----|
      160560
160570 TCGGTATCAGATGCACATATAGGAAGTTACTTGAACCATAGAT
...
160762 GCAAGCTGCAAGGACAAAGCTTCCTTTCCACGCTATATTTAAATATATATCGCCTGC
      |-----|
      160812
160822 TTTACAAGGGTAGAATTTTTTTTACCACACATTACTATAAGTGGA
...
160993 TGGATAAAAAAGCTTGGTGTCTCCTTGCTTGCTTAAAGTAGCTGCTATCTTAGCGCTATA
      |-----|
      161043
161053 TGACCTGATTTACCCATACCACATACTATGCAAGCGCCTTTGGTATTTAACATCAATTT
      |-----|
      161084
161113 ACAGCTTGATTAATAATTTTCATCTAAA
```

```
...
161282 GAAAAAATTTCTAAATTTCTCACTATCCACTTTTGGCAATACCAACACAGAAATCGCCTCGT
      |-----|
      161332
161342 TTCATATCTATACCACCTATGATATTTTTTCAGCTTTTTCATTTGCTTC
...
161598 CAATTTGACCCCTTAGGCTCTACTTGTAGCTCTTTGGATAAAATCTTAAATCGTTTTTG
      |-----|
      161648
161658 CTTCTTGAGCTCTTATAGCTGCACCTTGAGCTGTAAACAGAACCACT
...
161796 GCATGTTTGTATTTGGATTAGCTTTAATTAATTTAGAAATATCTTAAAGTCGTTCTT
      |-----|
      161846
161856 TACCCACAACTAAAGTTTTTAAACCAATTATGCTGCGCCATTTTACACT
...
162329 TACTGGATCACTGAATCCGAACCTAAGTCTTGTATAAAGTCTGTGGAGCATTAACGTTTCCT
      |-----|
      162379
162389 GTTACAAAAAACAACACTGACATTTTGTATAAGCATTCATATCTGCATAAA
...
162440 ACATTTGCTTTTATAGATTTTATTAAACACGTGCTTTTATAACACTTACAAGGCGCTATTT
      |-----|
      162490
162500 TTCACACCTAGTAAATTTACAGCTCCTACTTTAGGGATAAAAAATAT
...
162638 GGGTTATAAACTCTTTTGATTGTAATTTTAAATTTTCCATTTAAAAAGATGAGCGCCAAAT
      |-----|
      162688
162698 ACTAAAAATTTGATGAATTTTTTGTGAAGATTAATTTCTTTTATTGCTA
...
163095 TTTTGGCAAAAGTAGCTTCTATAGTTAAATCCTGAAATTTAGCAGCCAATCGTTAAGC
      |-----|
      163145
163155 TTTTGAGAAGAATTTGTAGCTGAAATTTTAGATTTTTCCTTAAAAAGCTGTTTTTAATC
      |--
      163212
163215 GCTTCAATTTCTGCTTTAAGTGTACTACTTTCAGTGTCTGCATCATATATA
      --|
      163217
...
163367 AACTCCATATTTATTTTGAAAGCAATCAAAATCATTTTGTGCTTTTGTATAAGCTTCTTT
      |-----|
      163417
163427 GTATTTTGCACCTTCATTTTTCAGCAAAAGCCATTTGCTCTCTAGCT
...
      |-----|
```

...
140483 TAACCTTTACACATTATAGCTCCAGCTATTTTGGCTGCACCAAAAGATAAGCGCCATAT
140533 |-----|
140543 TCTAAATCACCTTGAATAAATATTCCTCTTTCAATGGGTTTTAG
...
140660 ATACATTGTTTTGAATCATCAAAATTTTCATAAATAAAATCAGCCCTTTGAGCGACATTT
140710 |-----|
140720 GCAGCACACAAGGTAAGACTGCAGTCAAAAAACATATCTGGCT
...
142641 TATGGCATCCTTATATATATCCCATAAAGATTAGAAATTTTAATAGCAGCAACGCCAAGT
142691 |-----|
142701 TTATAAGCTAAATGATTAGCAATCTTTTACAGCACCTATTTCTTGATCGCTTTTAGA
142748 |-----|
142761 AGATCTTTAAAAATAAACATTTTAAATTCATCAAAATATTATAG
...
143203 CAATATTTCAATATCAATATATTTTAAATAATCTCATATGTTCAATGAGCGCAAAAT
143253 |-----|
143263 GATCAATATATTGATAAAAAATTTTATTTCTATCCAACTAGCACA
...
145544 TTTGTAGAAATTTCCAAATTTGAGGGCTTAAGATAATCCCCACCCCAATATATCGCCGCAA
145594 |-----|
145604 TGAATCCATTTTTTTGGATCAACTTTTTCTTAAATTTAATATAGCCT
...
146711 GTCCCTAAAGCTAANAATTTTAATTTATCTGACITTTATCTGTTTAAAAAATCGTCATGC
146761 |-----|
146771 TTTTCATAAAAAACATCATCTACATAATATGTAAAAACCTCTACAT
...
147247 TATCATAATACACTTTCAAAATTAATAATTCATGTTTTGTTTATTTGCTAGAACGCCATTT
147297 |-----|
147307 TCCAAATTTCTGCCCATTTGTGATACCTTGCATATCTACAAAAAGGA
...
148309 TTTTTCATGTTGCTTGAAATTTTTTATGATAGACAAAAACCATTTTTTAAATCGCTGCAT
148359 |-----|
148369 TAAACATCCTTTTAAACAAACTGCCAAAAATATTATATCATCTTA
...
148900 CAACTTATATCTTTTCATCGATTTCTTTTAAGTAAACATATAAATTTCCAAAAACGATGCAA
|-----|

148950
148960 AGCACATATTGGAGCAAGTGTTTTTTCAAGATCAGGATGTTTTTTTT
...
149843 TTACACATTCATAAATCAATTTGGATTCCAACTAATGATAATAAAGGTATATCGGTTACA
149893 |-----|
149903 CAATCACTATAGGCATAAGAATTTTTTAAATCAAAATTTGTTTTAAGT
...
150068 TAAATCTAGCAAAATCTTTAATAATAAATCGTAAGACCACCTGATACATATAACGATTTG
150118 |-----|
150128 TGGCCTTCATCTTGATGCCAAAAAAGTCTATCCATAACATTTTGAT
...
150315 TTTGCGTATAATTAATATTTTATAGATCCAGCTAAAGGCAAAATATCTATCAACGCTTGAA
150365 |-----|
150375 AATTAGTTAAAGTTTCACAAAAATCAAAATAAGGCCAAAAATTTTCCTT
...
152721 GGAGTAAAGTTTTTGAATATTTTTTATTTGCTTGCAATATCCAAACCTTTTCAAAACGCCACAA
152771 |-----|
152781 ATCCCCCAATTTCTTATCCCACTCTTTAGGAATAACTTTTTTTCACCA
...
153403 ATTTTGGGAAGATAATTTGCATTTATTTGTTTCAAAAAATTTTATCTGTAAATCGCTATA
153453 |-----|
153463 TTATTAAATAAATAAGATTTTTTCATTTAATATATCGATTTTATCCCACTATATTTCATGA
|-----|
153523 TTATCAGACAAAAGCGCTAGCTACATAAAATCTCATCTTAAATTTTATCTGCAAGAAAAAGAA
153534 |-----|
153583 TATATAC
...
154242 CAACCTTATATCTTTCATCGATTTCTTAAAGTAAACATATAATTTCCAAAAACGATGCAA
154292 |-----|
154302 AGCACATATTGGAGCAAGTGTTTTTTCAAGATCAGGATGTTTTTTTT
...
156266 AAAAAATTTAAATAAATAAAGCTCCATAAAAAATTTGTTAAATGGACATAGCGATAATT
156316 |-----|
156326 TTTTATACGTGTTTTTAATATTTTTTTTTTTTATTTGAAATTTTGTATACACAA
...
157469 TACATCTGCTGAAAAAACTTTTATCTTTAATCTCTTTAAGCTTATCATCAAAACGATTTGC
157519 |-----|
157529 TATGATCAATCAACTTTAGTCCAAAAACATATAAATCATCTCT

129412 ATACGCCAGATATATCCGTAATGATTCTTTATATAAACTAAAGCTAAACGCTTAGG
129462
129472 TCTTAACCTATGATTTTCAGCTGGATTCCAATCAGGCATTAACCCA
...
129520 AATAGCTTTATCTCCCAAACTCTAGATCTTGATTCTTTTAAAGACTCAAAACGTTTATA
129570
129580 AAATCTATATAAGGCTCTTTTGGTAAAGAATGAATAAATTATTT
...
130573 TAATTTGGTACAAATTCACAATAGTATCTTCAGCTATTTTTTGCACAGGATTCGTTAAA
130623
130633 TCCCTGTGATTCACATACAGAAGAAAATATCCCTCTGATTTTAATAAGCGATCGCCTCC
130679
130693 TCTAAGGCTTCATTTCTATCCATTACATTAAACTACTACCA
...
130876 CCTCTTCTAAAGCATTAATGCTTAAATTTCCAGTTCCTGCACCGATATCAGCGACTTTA
130926
130936 ATATCTTTTTTCCAACTAAGCTAATAAGCATATCTATAGTTTTTG
...
131215 TTCACCTGCTGTGCTTGTATATCATTTACAGAATGCCAAGCTCCAAGATAACGCTCTTT
131265
131275 ATCCCAAACTCTTTATAGTCACACTCCATAAAAAGCAGTCTATG
...
131618 GAATATTTGTATAAATTTTGTGTAAATTTTGGCCCTCTTCACGCATATATATCTTTGGT
131668
131678 TCCACAGCACTAAATTTGGCATCCAAACATTTTCTCCCAACATCTTAG
...
131974 GCACATTTTTCAAATTTCTGGTAATATTTCCAAGAAGCTACTATATCTGGACGATGTTG
132024
132034 TTTAATATAACCAATCTTCATCCAGATATTTTACCACCTTACT
...
132428 AATAGAAAATATAAATTTCTTTTAAAGTAACTTTTTGTACAAACTCTTTATCGTTAATC
132478
132488 TTAATTTGTTCAAATTAATTTAATTTGATAGATAAATTAATGTGT
...
132573 AAAACAATCATTAATTTATCAAAACATTCGATTAAGATTTCCATAAANAACGATTGCG
132623
132633 GCCTTTGTCTTCCAAAGTATTAGGCAAAATTAATAATTATATACCTC

...
133897 ACACATTCTTTCAAATTTCTTGGTAGTATTTCCAAGAAGCTACTATATCTGGACGATGTTG
133947
133957 CTTGATGTAACCAATTCCTCATCTAAATATFCCTTTATAAGATTTTGTGTACACTTATATT
134017 CTTTTTAAATTAATCCAATACATCGTCTACCTTCTTCTTTTATTTTTTTCAACATTAAT
134038
134077 CCTTTTTTTTAAAGAAA
...
134833 ATCATTTGCATGCACAAGCTTCACATAATCAAAAAATTAATAATTAATAATACGCTTACC
134883
134893 ATCCTCATATCCAAATGCACCAATAAGCACACCAACTTCTTTT
...
137177 GCTCCGTTATCTATAGCAAGTTTATAAATTTCTTTCAAGCTCATCTACTTAACTGATTTCT
137227
137237 GAAATGATTTTTTTTGACCGCCAAGATTTTCTCTAAAAATTTGGGCA
...
137564 GCTCCACCTACTATGCTTAAATTTCTCTCTTTCTATCTCGTAAGCAAGTTTAGCGATTTCA
137614
137624 TATTCTCTAAAGGTAGATTAAAGCATTCCTGCAAGGCTTTTTTATAA
...
138542 AATAATATCAGTAATTAAGCCTGCATGTCCACCACAAATCAATACATCTGGACGATATCC
138592
138602 ATCCCCCCCCCATATTTATCGACATAAAGCATCTTGTA AAAATA
...
138893 TTGTTCTTTTTTCATTTATTTATCTCTTTTAAATTTTAAATTTGCAAGCGATTAAC
138943
138953 TTTTCATCATTTTAAACGCTCTTAAAGCTCTTTAATACCATCTTCTTAAAGTATTATCTGG
138966
139013 TTTCCAACC
...
139759 ACATCAAAATTAATCAATCACACAAACTTCATTAACCTTTTCAAGTAATAATTCGGGTTAAAA
139809
139819 CCGAACCTATATAACCTGCACCACCTGTAAATTA AAAACTTTTTTTTGA
...
140147 AGATCTTTTAAATCAATATTTTGCATCGTATAGATAGCTACATCGGTAGATCTGCTACTA
140197
140207 TGAATAAATTTCTCTCTAACTTTTCCAGTGCCCCCAAAATTTCTACAC

119274 GCTCAAGATATGTAATAATGCACCTATAGAACTAGAACAAAAATAGCGA
...
119377 TACATTACAGCTAAACCTTTTGAAGAAGTAAAGTACACAAAAAGCAAGACGTTTAAA
119427
119437 AGATGAGCTAGTTGGAAAAATCAATGAAATTTTAAACCGATGGCTTT
...
119492 GTTATTTTACGGATTTTGTAGTTCTTTAAATGCCGTGTAGGCTGTGATATCATCGCCATTT
119542
119552 CTAGATAGAAAAAATTCATTCTAGACATGGCAAAAAATTTTGTGA
...
119987 ACAATCCTTGCCCAATTTTTCGGAATTTCTTTCAGCCACGCATGACAAGAAACGCTCTCC
120037
120047 CAAGCATAGCCATTAAACAAGCTTGCAATTTACAAAAATGATCTAAGG
...
120196 AGTACCTGCTAATTTTCTTTTAAAGCTTCCTTTAAAGTGGCTTTTACACGACGACGAAA
120246
120256 ATTGGCGGAATTTTGTAGTTCTAACTCTTTTCATAGCTTTCTTTTGTAAATTTTATCCTC
120316 ATACCATTAGCCGACGACGTGTTAAAGCCCTACTACATCAGGTACAAAACCTTCTTCA
120341
120376 AGTTCTAGTTGTAGTTTTTAAAAATATTAGCCAAAATCAAAATCGCTTCTTGCATAGCA
120418
120436 AAACCTTGCTCTATACAAATTTCTTCCCTACCCGAAA
...
120758 TCTCCTCGTAAGCTTTTCTTGTCTCCTTAGGTAAGACTAAGCAAAATAAAGCGTCCATG
120808
120818 TCAAAGAACTTGCTGTGTTTCATGTCTGCTAATAAAAGCATAGCCACTTGATCTAAAA
120878 TTTCTCAAAAGAAAACGCTTATTGTTATCAGCATCCACTACTAAAAGCAAGATCCTA
120893
120938 AAATATCTCA
...
121206 AGTCATAGCTTCATCTACTTCGTATACAGCATGATTTGGATACCTGTGTAACAGATCCAT
121256
121266 CATATCTGCCACAGCTTCACTCATAAAGATTAAACACTTTTATTGATA
...
123270 TCCTCTATAAAGTGCTCTATGCTTTTGTGTAATGATATTTAAATTTGGGTATCGCTATAA
123320

123330 ACTTGTAATAATGCCCCAAAAAGTACTTTTTTCTATAAATATAAAATCTTTTCATTTAGTTTT
123339
123390 TTTTC
...
124340 TCAAAACCACTTGTAGAAGTAATAGTGTGAATTTTCATCAGCAATTTTATAGCGTATCA
124390
124400 ATACTGCAATCTTTTACAATCTCATCACAAAAATTCACAAGATGAGTG
...
124901 AAATGGAGCAAAATCTCATCTTGGCCTTTTATATAAGAGTAGTTTTAAGACGTTATAA
124951
124961 TCTATCCTTTTACCCCAAAATGAAAAATTTATCATTTTCTTTAAGTT
...
126325 TAAATCTCATTAACAATTTGAAATCGTAGTACAAATAACCATCATACCTTGATCGTTTAAA
126375
126385 AATTTTGCAAAATTTAGCTCTTTTCAAAGCCATATCTATCGGACCTT
...
126868 ATAACCAATTTGTCACTCTTTAAATGTTTTCAGCATCTTCTAAAGGATTTGCAAGCGTTTGT
126918
126928 CCAAGCTTACACCAAGACTTATCAACTACTATAGCAAGTTCTTCT
...
127124 TTAATAAAAAATGCTCAATATCATATATTTATTAAGATAATTTTTAGCACATCGTTTAAA
127174
127184 TATCCACCCACAACAGCTATTTCAATTTATACCTGTCTGATTTTAAAG
...
127863 AATAATTTTCCCAATCCAAAGCCAAACATTCCTCTTTTTCATGATAACTATCGTTGCAA
127913
127923 ATAAGCCTTTGAGTAATACCTATAAAACAAATATTTTGTCTCACTTTC
...
128661 TATCTTTTGTGATCTGCATCAAAATAAGCTCAAAATCTTTCATCATACCTTTGGCGATAAAA
128711
128721 TATTATAAGTTCTCTCTTAATGTCCAANAATCTTTTAAAAAATTG
...
128899 CATTGCTACAAATGCAGCCCTAGCCACCCAGCAAAAGCAAGGTTCCATACGCTTGA
128949
128959 TTCTTCAATAAGCCAAATAAATTTTATCAATTTAAGAAAAAATTAGAA
...
|-----|

113327 CCATAAGCAAGGTACTTTGGTTAAACTCGCTAAAAATTCAGCAGCACTTTAGCGATATCT
|-----|
113377

113387 GTACTTTGTGGCTCAGCTCCGTTCCCTATAGAATAAATAGGCTCATAGCGATGATGAGT
|-----|
113434

113447 TTTTCATAATTAAATCAATAATTCAAAGTTGTTTTTTAAAAA
...

113723 CCTTGTATGAAGTTATTTTCATTTTCTAAAAATGCTATACTAGGTGGAAAAACGATAATA
|-----|
113773

113783 TCATCGCATTTTACCCCATAGTTTTTATTTAAAAATTTTAGCATAAA
...

114080 CTTTGTGAAAATTTATCAATTTCAAAAAACACCCATAGTCCATCCACCAAAATCGTTTGGC
|-----|
114130

114140 CATCAGAAATCACCTCCCTTAAAAAGTCTCACACTCGCAGGTCCTAT
...

114445 AAATTTGTCGAATGCTTGAAGTTTTCCTACTTACCTTTGAACCTCCAACAACAGCGACAAAA
|-----|
114495

114505 GGACGTGCTGGATGTTTGATGAAGATTGCTTGCAAAAATCAATTTCTT
...

114551 TTTGGAGTAAAAATCCTCGCGCTTGTGTTTTTTCATCAAAAAATTTAGTAATCGCCTCTA
|-----|
114601

114611 CACTAGAAATGTCCTATGGCATACTCCAAAAGCATCATTTGTATATA
...

114789 GCCGATACATCTTTTGCCTAACAATTTCTTTATCTAAAAGTCTAGCAGACGCTTAGC
|-----|
114839

114849 CACAGGTTCTAAAGAATATTTAGAGCTAATTTCTTTTGGAGCGCCCTAAATGAGAGGCTAA
|-----|
114887

114909 AATCACACTACAACCATTAATCCAAAACAAATATCTT
...

115070 TATCACTCATGATTAAAGCCTTATTTGCAATATATACTTGCCATATCTACTAGACGACTTGA
|-----|
115120

115130 ATAGCCCCATTCATTATCATACCAAGCAATTACCTTGATATAAAATCA
...

115185 ACTTGAGTTAAATCACTTGTACAAATAGTCCATAGAACAAGTGATAAAATCGCTTGAA
|-----|
115235

115245 ACTCTTTCATCATCATCTACCAATTAANAATGCCCTTTTAAATTTGTAG
...

115448 TGCTGCGCCTCTAGAAAGTTGGAATGATATTTTGGGCTGCAGCAGGAGAACGCGCTTATC
|-----|

115508 TTTAGCTTTAGCATCAATAACTACTTTGTCCATTTGTATAGCATGT
115498

...

115936 TATCGTATTTAAAAAGGTATTTTGTAGTTCAATATCGTAGTATCATTTATCGCCACAA
|-----|
115986

115996 GCTCAATATCATTTCTTCTTAAGATGATTTCTTGAACACATCTGCC
...

116143 AGATAGCACATTTTGGTGGCAGTTTGTATCCACCGCATAAATGGTCATAAATAGCGTTGPTT
|-----|
116193

116203 TAGAGGCTTTGGAAAAATTAGATATTTGATAAACCTTATTATTATGCC
...

116321 TTGGGACATTTGGCCAAAAGTTGAAATTTGCGATTTTGAAATCAGGCCAAAACGCCCTGT
|-----|
116371

116381 TCCTAGCATAGAAAAGTGTAAAGTATCTTTATAAGCTTTATATATCCT
...

116533 ATGATATAGGAATTCCTAAGAAATTTCAAGATTTTAAAAACCAATAAAAAATCGCTTCTT
|-----|
116583

116593 CTTTCATACGCGACACTTTTAAATACAAATAGAGTTTGTGAAGAAAT
...

116683 GATACAAATTTCAAAAAATTTAAGGACAATAATGCAAGAAGAAATAGATTTAATCGTTAAAA
|-----|
116733

116743 TTTTAGATGAAAAAAAAGCTGAAAAATATCAAAAACCATTTGACATGAG
...

117585 CGTGTGATAAGATAAGGCTGGTATTAGATGCAGGAAGTAAAGCCCATCCATCGCCAAA
|-----|
117635

117645 TCTATCTAGCTCCATCAATTTACAAAGACTTTTAAACACCTTTTA
...

118743 TTACAACCCCTATTTAACATAGTTTGTCCCAACAAAAACCTATAGCTTTAAGCTTTT
|-----|
118793

118803 TCATTCAATCTTTTGTGTAAGTCTCTTTATATCATATTTCTCTAA
...

118853 CACATCTAACATTTTTTCTTCAAAATCATCTCTTCTTTGACCTAAATTTATCTTTTC
|-----|
118903

118913 TAAAGCAAATTTGGATAGAAAAAGAAATTTTGTCTACAATTTTACAT
...

119214 CTAATATAGGTCTATGTATCCACTTGATCCTTTTACCCCTAAATTTGTGTTAGCGATAGCG
|-----|
119264

105265 ATAGAAGAAAAAATCAA
...
105407 CTAGTGGATATTTTCTAAATCCTTTAGAACCTGGAGTAGACATTTTCAGTTATCGCCGCTA |-----|
105457
105467 CAAAATATCTTAGTGGACACTCTGATGTAACCATGGGTATAGTACT
...
106506 TTTACCGATCACCTTTGGGCATAAGGTAGCCATCTGGCTTAGCGAGCTTGGATCGTCATTT |-----|
106556
106566 TGAAGAATACCTTGAACCTTGA AAAAGACTTTATGTAGGAGCATAGTAGGAGCTGTGG
106626 AACCAAGCGAGTTTAAAGCGATGAATGCAACGAAAGTAGAAAAAGCTAACTTTTAGAGAAATTT |-----|
106642
106686 AGGACTTGAAGTGCCTAATATCTCTTGGCAGAGCACACGCTGATCGCTTTATAGAGCTTGG |-----|
106727
106746 TTTTGTAAGCAATATCAATGCAACTTTTACAAAAATCGCTCATCAATTTACTCATACT |-----|
106783
106806 CTCACAAATGAATCGATGAAGTAGGAGAACC
...
106901 AAAATGCAGTTACAGTAAGCAATGCCCTTTAAGGCAAAATTTAGCCATTTTAAAGCGCATATGG |-----|
106951
106961 AAAGACACGAGCATGAAGAGATGGGCAAGTTTGGAAAAATGGAAATG
...
107111 TTGACACTCTTAAAGGCTTTGTACTGCGAGAAGCGGTTATGTTCGACTAAGCGCATCTT |-----|
107161
107171 ACGGCAAAACACACGCTCATGAATCGTTTATGAAAAACGCTATGCTAGGTATCGAACAAC |-----|
107194 107206
107231 AAAAAACTTTCAAAGAAGTATTGCTTGCAGA
...
107327 TAGGATATGCACCAAAAACTTTGTAGATGAATTTTGGCAAAAAATCAAAAATAGCGCTATCT |-----|
107377
107387 TAAATAGGAAAAATCATGTTCTATAGCGAAATTTTGGCGGATTTT
...
107995 CTAATTCAAAAGTCTTACATATAGCAAAATGCCATTA AAAATGGAATTTTAAATCGCTAACT |-----|
108045
108055 TTGCTAAAAATAACATGAGTGAACCTCTAAGCATTTTTTAAAGGTAG
...
108826 TGACTTTAAACGAGTTTAAAGATGGACAAAAGGCTATTATTGTCAATTTTAAAGGCTCACA |-----|

108886 AAGAACTCAAAAAATAGACTTTTAAAGCTTTGGATTTTATTA AAAATAA |-----|
108876
...
109797 GCTTTAATACCTTAGGTGAATTTGTTAAAAACAATATTCTTAATACCTTTTATCGCTTCAG |-----|
109847
109857 CTTTGGCTGATGGGATTATCGCAGGTGTTGGTGTGTGATACTTTT
...
110634 ATTCACCTAGAAAAAGATGTAGATGAACAAATAATGATTTTAAAGGAATTTATCGCTCAAAA |-----|
110684
110694 ACATACCTATACCAAGTGCAGTGGCTTTTATTTATTTTGTATGAT
...
110861 CATATTGCTAAAAATTTTATTAAATTAATTAATGATGAAATTTTATCGGATATTA |-----|
110911
110921 TCCTCAAAAGCTTTTAAATGCTATAAGCAACACCTTCAAAAATCAA
...
110981 TAGATTAAAGTCATACGA AAAAGATTTACTCACATCAAGCTTTGGTTTTCAGCGTTCATA |-----|
111031
111041 TTTAACTTAAATCATAAAGACTTGCATGTTTAAAAAGCGTAATTTT
...
111548 GGTCCTTTACAACCTATCTTTAAATAAACAAGTCCATTTCTTTGCAAGCTTCATCGTCAAAA |-----|
111598
111608 TGTTACGCAAAAATCCCGCTTTCAAATTTCCGCTTTCTTTCCACATATCATGCAAAACAT
111668 TTACCATGAAGCCACACAGAGCAGGATTTTGTCTCATCAAGTTTCAGGCAAAACTCCAAATAAA |-----|
111687
111728 GCAAAAAAACTTAAA
...
112388 TTTTGATTATCCCAACAAAGAACAGTTTGGCTTCTTCAATTTTTCACATCGCCCATTA |-----|
112438
112448 CCCATGATGTTATATCCTGTCATCAACATAATGTATTTACCTGTGAACACCATTTGCTTAAA |-----|
112507
112508 TCAGTAAATAATACATAGCCGAATTTACCTACATCTTCTATACTAACATTTGCGTTT |-----|
112512
...
112934 ACCGCATGCACACTACAAAATCAATCTCGCCCTAAATCTTTTGTGATTTTTCAGCGATAGAA |-----|
112984
112994 TCTAAATGCTCGTTATTATTATTACATCAAGCTCATAAACAAAATTAG
...

98428 TTCTCAAAAACAAGTCTTCTTGCAGAACAGCTTATCTTGGAAATAAAGACGCTGTCA |-----|
98478

98488 ATTATCTTTAGAAAAATAAAATTATAAATCAAATAGTAGTTTTTTT

...

98651 TGATTATCATTTGTCAGAGCAAGTCCACAATAATTTACCATCTACACCGCATCGCTTGCT |-----|
98701

98711 TCAAAATGTATAACCATCAGTTGAAACTTCTCCTACCAAAATTTGCAC

...

99616 AAACCTCTTGCTATTTTTTGCATATAAAGAGGAATTTTTTAGAGCACTATCAACGCTCGCA |-----|
99666

99676 TTATATTGAGTTAAAAAAGTCAACATAAAGCTCTTGATTGTTTAAATAGCGATAAAATTCG |-----|
99722

99736 GGTATAAATTCCTGCTTATCATCAAAATCCAAAATAACATCTTT

...

100226 ACTAATTAATAAAATTTATTATTATTAGGAGAAAAACAATGAAAAATTTGACTAAGCATTTTG |-----|
100276

100286 GAAACAATTATAGCTGATACCAAAAAATTCATTTAAGCCGAGCGCAAA

...

100739 ATTTTTGGAGTTTATGTCTCTGAAAGTTTACATCAAGTAACCATCTTATGACGGATAGAG |-----|
100789

100799 GAATTCCTGCAAGTTATCTGTCATATGCATGGATTTTGGAAAGCCACTTATAGTTTATTTA |-----|
100814

100859 ATCATAAAAAT

...

101153 CTCAAAAATTATTTTAATGAAGTTGAACAAGCTGCCTTTTAGTCCAGCAATATCGTTCCCTG |-----|
101203

101213 GAATTGGCTTTAGCCCTGATAAAATGTTGCAAGCTAGAAATTTTTTC

...

101297 ATCATCTTTTACCAGTAATTCGTGCAAAAAGCGAAGTGAATACTTACAATGTGCGTGGTG |-----|
101347

101357 CTATGAATTTTGATAGTTATATAAAATGATGCTGCTTATTATGAACC

...

101522 CTTTATTTAATCTTATGAATGATGATCAAAAAAAGTCAACTTTTTCATAATATCGCGCGTT |-----|
101572

101582 CTATGGAAGGAGTTGATGAAAAAATTTATCACTAGAGCTTTAAACA

...

101833 TTAAATCATGATAGAACAGGTTTAAATGCTAAATTTAAAAACTCAATAAGCGCATAGCT |-----|
101883

101893 TGCTTATGCTTGCAGCTTATAATAATCTTATGAAAGCGCTAAAAATGCTTTTAGAAAAAG |-----|
101928

101953 GTCCAAAAGTAGATGAAAAAACGATCGTGGACAAACCCCATTAGCGGGAGTTTGTTTTA |-----|
101973

102013 AAGGATACTTACCTAT

...

103332 CTTTTCGGGTAAAGATCTACCAAAAAAGCTAAAGTAGAGATAGATCATCGCTCAAA |-----|
103382

103392 AAGGATAAAACATGCTAGGTATTGTTTTTCTTTTATTAGTGTTTT

...

103497 GTTTACTTTTACTTTCCTTAGCTGCTCTTTTGGACTTCTCCACATATTGTGCTCAAAAG |-----|
103547

103557 GTTCTTTAAATTAGGTTTTTTTGATATTTTCAAGTTTTTCAATCA

...

104307 TGGTGATTTCAACTTTTGGGCTATTATTTTGAAATGATACGCTATAAAGCGTTGTAG |-----|
104357

104367 AAACCTCTTAATCTATTATGATTTTCTTTAAAGGAATGGACATTTATTGTTCACAG

104427 TTTCACTCATCGTTTTCGGTCAAGTATTTGCAAGTGGTCTTTTATCTGTAGGTTTTGTAG |-----|
104435

104487 ATACTTTTAATAGAAATTTTGCAAAAATGCAGGATTTGGAGTTTGGCTATTATTATCGCCG |-----|
104540

104547 TTTCAATCTTGCCTTGGGTTTGTGCTTCTTAATGGGTCAGGAATGC

...

104672 CCAATTCAATCATGACGGGCTTTGGAAGATGTGTAGTCCTATTGCACCAGCGATCTTA |-----|
104722

104732 GCTATCTCAGCTATTGCAAAAAGTAAGTCCCTTTTGGCGTAGTTAAAA

...

104841 AAAAAGGAGAAAAAATGAACCTCAAAAACTAAACTTATTCATGTGCAGAGCGCATCAAG |-----|
104891

104901 GTGCTGAAGTAAGATCAGTCAATCCCTACACTTATGCGTGCATCAAC

...

105145 AGCAAAAGATGCACATTTTTTTTAACTACTGATGCAATTTATGGGCTCTAAGAACGATTTGC |-----|
105195

105205 GAATTATTTTAGACAAAAATAGCGTTTGAATTTGATTTTTTAAAGCTGATGCAAGTGAT |-----|
105226

```
...
90161 GATTTTATAAAATTTATCCTTGTGCGTAGCTTTAAATACAAAATGGTAAACGATGTTA |-----|
90221 ATCAATGCTTAATCCATTTTCATAGTAAATATTCCTAATTTAAATTT 90211
...
90963 TTCTAGTAAACATTTAGGCTTTTGCTTCTGTATAAACCTTTGGATTTGTTATAGCGCTTCTA |-----|
91023 TGGTAGCAGGTTTATCCCTATGGGCTTTGCTGGAAGTTTAGGGGAAAAAAGGTTTAG 91013
91083 CTAACAAAT |-----|
...
91406 TTTTCTTTAAAAAATCCAGCTGATTTTAAAAACCTTTATCTACCTTTTAGAAACGCTTTTA |-----|
91466 AATTTCAGTACTTTTGGTGAATTTAGGGGTTGTAGATATTTATA 91456
...
92661 CAAGTCGTACAAAGAGAAGCATTTAAATTTCAAAGCATTTTCTTGTATTATCAGCGATTTAA |-----|
92721 GCTCGTACATTACAGGCAACCCACACAAAGAAAGACAAAGCGTACAGG 92711
...
93043 GAGATACAAAATCCGTATATAAAAACCAAAATTTGCTCCACTATTTTGCAAAAGCGTTAAAA |-----|
93103 AATGCATAGAAAGAAAGCCATTTTTCACCCCTCTATCATATAAAGGGTAA 93093
...
93226 AAAAATCACAAGGGTAAAGAAACTTTCTCTTGGCATAGGAGCATATATCAAGCGATCCGC |-----|
93286 TAGGACACAGCTAATACATCTCCACACGCCAAACAATCCACTTGAGAAAAATTTCCAAAT 93276
93346 GTTTGTTTTCATCATCTTTTAAAAATCCGCACAGTAGGACAAAATTTTCAGCACATTTTGCAC |-----|
93406 AATGCTCACTGCGTCTTTC 93369
...
93525 AAATTTAAAGATTATAACATCCACTTTGTCTGTAAATCTTTTACGCACAACGCTTAAA |-----|
93585 TCCTCAGCTTTAAAAAGTAAAAATCAAAATCTATTTCCACTTCTT 93575
...
93699 CTTAAAGCAATAACCTTTAAATTCATGTCTTTTAAAAAACHCTACAGTTTGAGCGCTCACCA |-----|
93749
```

```
93759 TCTGGGACAATCAAAACGATTTTGCACCTTCTTGTGCTTTGCTCTAAATCAAGCCCCAAATCA |-----|
93819 TAAACGCTAGCTCTTGTCTTCATAAAGCGTTTAAACATTTTACTTTGCTCTAAATTTTCA |-----|
93879 TCTTGAGAAATTTTAAAGATA 93821
...
94493 AAAAACCTTAAAAATGCTTTTTCGAATGTGAAGATGCTTTAGTGGTCAATTAACACGCTGCGG |-----|
94553 CAGTTTTTTTAGTGCTTAATTCCTCTTGTATGATAAAGAAAGTCAT 94543
...
95473 ATATTGATCATGAAAAAATTTCACTCATTAAGGCTTTTAAATGGTTTGAAGCGCATAGTT |-----|
95533 TAAAGAGAACAAGAAGACAAATCACCATTAAATTAAGTTTTTC 95523
...
95896 AAGGCATAGAAGAAATTAATAAATTTATCTTTATACCATAGAAAACAAACGATGAAG |-----|
95956 AACTGATTTTTCATTTATATATCGATAGAGTTTTTCTCTAAAAGGCATAGGAAGTCTCG |-----|
96016 TTACAGGCAGTCTTAAATGAAGGAAGCATCACTTTAAATGAAAAAATCATCTG 96012
...
97373 TTCAAAAGCATTAAGCACAAATTTCCAAATGCTACCGTTAGTGTATCAAAATCGTCAAAA |-----|
97433 GGTGTTGTAATACGGGTTTTTGAAAGTGTATTGTAGCGCTTCAATTAATGACAAAAACA |-----|
97493 AGAAAAACATTTCTTTTTTACAAAAGATAGTCT 97467
...
97913 AAAAGTAAAGTATCCAGAAATTAAGAAAGTTTTTTCCTTTATATGAAAAATATCGCTCAT |-----|
97973 AGGCTTTAGCGCTACTCCAACTATTATTAATAAACCTTAATGTCTAAATTTGCTTTAAT |-----|
98033 TTA 97980
...
98304 AAAAATCACAACCTTCAGCAAGATTTTTTATGATTTTGATCCTTTGGAAGCGATGGAG |-----|
98364 AATGGATGGTAGAAAACTACTATACTATGCAAGTTAATTTACAAT 98354
...
```

83227 TCCTAATCTTCATGTAAGCTATGACTGCTGCTGAAATTGATTTTATCATGTCGTTT
|-----|
|-----|
83277
83280

83287 TAAATGAGTTATGAAGAGTGATAGAAAAATGCTTGAATACGGCGTTGATTCATGCC
|-----|
83330

83347 AGTGTGGAGCTGAATTTTGGATGAAGAAGTGCCTAA

...

84159 ATGCTTATTTACCTTTGCGGTTTGTTTAGGTCAAACATAGCTTGGCAAGCGCCTTAG
|-----|
84209

84219 GAGCTGTTTTTATTTCAAGTATTATTTTCTCATACTTTCTTTTAC

...

84588 TTAGCATGGAANAATGGTTTAGGAGCCATCTTTTACAACTTGATATCAAAAGCGCCTAA
|-----|
84638

84648 ATATTACTATGATACCTATTAATCTTACTTTTTTATCACTCAACT

...

85800 ARGACAAAATTTAGATGATTTTGAACAAGATACCACAAACTATCAAGTAACGATCCAC
|-----|
85850

85860 TTTCAGGATATAACAAGCTATGACAAGCTTTAATGTGGCTTTGTATGATTATGTGTGC

85920 GTCCCGTTCTTAAGGCTATACGCTATCACCTCCTGAATTTATAGATTAGGGGCAAGAA
|-----|
85940

85980 AATTCTTTGATAATTT

...

86755 TTATAAAGATCAGGTTTTAAAGTCAAAAATGGTGAGTTAAAAATGAAAAAGCCTATTT
|-----|
86805

86815 TTTAACCACCTTCTATGCTAGTAGATGGGAAGAAAAAATATCATT

...

87025 GCAAAGCTTTAAAGTTTTTTTATTTCTCGTCCAAAAAGTACTTTTTTTGGAAGCTTGTTT
|-----|
87075

87085 ATATGCCTGTTTTTAAGCTTTTTTGTCTTTTAAGCTTAGCGTAGATG

...

87677 GAAATTATGAAAAAATAAAGATACGCTCTACCTTGGTGGCGTAAGCATGATCGTGATG
|-----|
87727

87737 ATATGATAGCTTATATAAATCTGATCTTGTGCTTTATGGTGTCTCTGTGATTTTTTAC

87797 TAGGACTTGCACCTTATTATTTTTTCCGCTCATGGCGCTTGTGTTTTTTACCTCTTTTA
|-----|
87830

87857 TTTGTTTTATAAGCTTAAAGTGGCGTAGT

...

88042 CTTGAACCTTACTAGCAAGCAAAATCCTAGCCTTTATGCCATTGTTACAAGCATGATA
|-----|
88092

88102 GGATTTTTTCTTTTAAATCCTATCAAAATATCGAACCTATCATATAAAAC

...

88257 TAAATTTAATCTTTTAGCCTTTTTGCGCAAAACCGCTCTTTGATCCTAAAAAGCGTCGCAT
|-----|
88307

88317 GATTTATGAAATTTTACGTTCTTCTGCTATAATCTTTAGCTTTTAATAGGC

...

88454 TAGGTGGAACCTTACCTTTTAGAAGTGATCATTCGTTTTTCCAAATAATAAAAAAGCATCAAA
|-----|
88504

88514 ATGCAAGCGTACTCTTGATAGTTTTTGAAGTGAATTTGAAAATTTAGCCACACAAGAAA
|-----|
88519

88574 C

...

88646 TAGGTTTCGGTTTTAAGTCTTAATAGTCTTTTAACCCCTAGGGAANAATTTAACGATGGAA
|-----|
88696

88706 AAGAGTTGGATGATTTTGCCCTTTGCGCTTTTAAATGAAAAATTTGCC

...

89123 TTCCTTTGGATATGATGAGTATTACTATAGCTGCTATTTGCTATAGGCATAGGCGTTGATG
|-----|
89173

89183 ATGCTATACATTTATTTATCGTTTTTAGAGAGAAATTTAAAAACAAAGCCTAGAGAAG
|-----|
89201

89243 CCATTATGATTTCT

...

89524 TCCTTTATTTAAAAATCTTTAAATGATTTTGTGCAAGCTCTAAAGCTTTAAACGATGAGA
|-----|
89574

89584 GATATTATTCTTCTCATCCACGCTAAGCTGGGCTAAGGTTTTTATCA

...

89783 AAACCTTTTTTTATCATTTTATGACAAGTTTATACGGTTTACTTTTATCATCGCCCTTA
|-----|
89833

89843 CCGCTAAATCGACGAGAATAAATTCCTGGATTTCCTTCTTAAACATCTACACAAATGCCA

89903 CTATCATCATTTAAAGCGATAAAAATCTTTTTTTTGTCTTTTTCATCCAAAGCATTTAAAAAC
|-----|
89917

89963 GCTCTTGTCTTAATCAAGGCTTTTCTTTTAAAGTTTTTCCATTTTCTTCAATCTCAAAA
|-----|
89980

90023 GGCATTAAAACTT

74871 ATAAGCTCATCAAGCGCAGGATAAGCTTTTTCCTACTTGTCTCAACCGCAACGATACCTC |-----|
74921
74931 ATAGCTGTTGGAAAAAGTATCATTTTGAACCTTTTGACTTCATATTAAACAT
...
75381 TTTGATTAATAATTTTATTATTATAAGCTAAAAAATAAATACTTTCAAAAACGTTTTTCC |-----|
75431
75441 TACTCAACTCATTAATGCAATTCATAGGCATTATGAGAAAAATCTA
...
75888 TGAAGTTTGCAATGTTTTATATTATTTGCCTTCAATTGGCTGCTACTACAAAACGATTAAAG |-----|
75938
75948 AGTATTAAATCCTACACCTATACCTTAACCTTAATATTTTATA
...
76024 AAATCCATGTTATTAACCTCTTCAAAATCACTACCATTAATAAATAAATTTAGCGCTAGAA |-----|
76074
76084 AGAGATAAGCATAAAGCTAAAAAGCCTGATGAGAAAAATTTTCTCCTA
...
77204 ATACATCCTGATCACTGAGTCTGTTTGCACTAAAAATCACTCAGTCCCTTTTAAACGCTTTT |-----|
77254
77264 GTGCATCTAAATTTCCCTGTGCCAATAATTTCTTCTTCTCACTCT
...
78217 ACCCCAGCTCCATAAAATTTTGGCCCCATAAGCCACTCCATAAGTTTTCCTATCGCTATC |-----|
78267
78277 TTTGCTCCAACAGCTATACCTGCTACATGACTTCGGTGTGTGATT
...
78577 CCAAAATTTCTTGGCATATCCACATCATTGGCCAAGTCTTTATGGAATAATTTCCATCGCCAAA |-----|
78627
78637 GTTGAGTAATGACCATCATTTTCAAAAAATTCACACATATAATGATC
...
78941 AGGATAAAACACATCTCTGCTATATAAGAAAAAAGCCATGTCCACTAAATAACGCTTGGAA |-----|
78991
79001 TAAAGGATGAATTTTTCATCTAAAGCCTGTTTTACACTCAGGCAATTTGGAGTATGC |-----|
79003
...
79275 GCAATCTTTAAAGCCTCTAAAGTGTCTGCTGTTTCTCCACCTTGAGAAATAACGATAAAT |-----|
79325
79335 AAAGAACTCTTTTTTATAATTGCTTCACGGTATCGAAATTCACATTG

...
79470 AGCGTGATTTTCATCTACTTGACTTAAATCTTCAATTGTTTAGCTCATAAAAACGACTTCA |-----|
79520
79529
...
79530 TCGCCTTGAATAGCCCCCATTTAAACTTCACTCATCAGCGACTTTGTTCATAAAATTTCT |-----|
79534
79590 TTTTCCATAAAAAACGAAAGCCGCTTTTTTTTGCATAAGCCTTATCGCCTGAAGCTTA |-----|
79634
79650 GAAAAACAAAGTGATTTTAAAAATATCATTTTCAAAAGCAAGAGCTCATCTTTACTCGCG |-----|
79685
79710 TATCCCAAGCTTAAATCTTCAAGATACATTA
...
81313 CAATACGCTCTATATATGCTTCTTAAAGAAATTTTCACTTTTCACTTTCTATAGCGCTTAAA |-----|
81363
81373 AGTATCACAAAGCTTTTAAATCATTTTTTTCATCGGTTTTTAAGTTTTT
...
81454 TCACCTTAACATAAGAAAAAATTTGCTCATCAAAACATATAATTTCCCTTTCATCGCTTAAA |-----|
81504
81514 AAACGAAGCTGCCAAAAAGGCTTACATAAATCTTTTAAAAATTTCCAC
...
81920 GAACATTAATAAGTAAAAAATTCCTCATCTTTTAGCAGCTTAATAAATTCAGCTGATT |-----|
81970
81980 TTAAAGAAGTAAAAAATCTCCAGCTAAACGCACTCGCTTAATTT
...
82565 CATAGCTTTTTTTTATGAAATTAATCCATATTTTTTAGAATGATATAAGCTAAACGCCCTT |-----|
82615
82625 GCTTAAGAAAAAATTTTCTACAACCTTTGGGTATAGGAACATAAGATT
...
82948 TCACAAAAACGCACTAAACTTCATGGAAAAAAGCTCTATTTTAAATATCAATCGCCATAT |-----|
82998
83008 CAACCCAACTAATATCTGTGCAGATACCTGTAAATTTTTCGCGCTTT
...
83061 ACCGAAAAATCCTAATCCTTATTGTGACTCATGAAGAAATCATGAAATCGTCGATG |-----|
83111
83121 AAACCGTAAACGAGAAACCAAGAGTGCATATCGTTTTCAGCTCACAAACAAGATACAA |-----|
83153
83181 CTTGCAATGCTATTTAGAAATTTTAA
...
...

...
69023 GCGTCGCAGGCTTAAACAACACTTTTTTCAGCTGGAGTTAATGCAAGCAGTAGCGATGAAA
69073
69083 TGACCCCGTTACACAATAATAAAGCCATAGCCAAATATTTACACCAAGAAAAAGAACTTT
69143 TGGAGCGTTATTTTAATGAATCACAAGTGAATTAGAAAAAGAAATCTTTTTCATAAAA
69146
...
69962 ATAGAGTTGAAAAAATTCCAATATAGTTCAGCTTTAATTGATGCTGCAAGCGATGAA
70012
70022 AACAACTAAGTGTGATGGTAGAATTAAGCAGCGCTTTGATGAAGA
...
70281 TGCAAGAGATACGACAAGCTTTTTTCACATCTTATCGGGCTTTAGCAAAATCGTCGCCT
70331
70334
70341 TCAAACTCTTTCAATGAGCCCAATCAAAATCAAGAAAAAGTTTAGAA
...
70423 GGTGTGATTAGCTAAATGAATTCCTTTGTAGATAGTATATCAACAGCGCTTTAT
70473
70483 GAAGCAAGTATGGAAGGAGTGCAATCGATCTTATCATACGCGAA
...
70632 TGAGCCAAATTAATTTATATCGAGTGCAGATTGGATGCCAAGAAATTTAGAGCGTCGCTT
70682
70685
70692 AGAACTTATGACTCCTATATATGATGAAAGAAGTAAAGCTAAATTAGCT
...
71125 TCGCTCATCTTCCAAAACCTCAAAAGCAAGCTAAGACTTTATTTCACCTCAAGCGTTGTC
71175
71185 AAAGCAAGGCTACACTATCAAAATGACCCGAAATTAAGCTAGAT
...
71245 TTCAGGCTCAACTGAAGCTTACGCAGCTTTTAGCGTAAGCAGGAGCGAATTTAAGCTTTT
71295
71305 GCCTTTAATTTTATTTGGGCTTTTACGCTTTGCCCAAGCGACATGCCATCTAAGCAGA
71343
71365 CTTACTCCTGTCGAAGCAAGTCGCTCCCTTATTTAGAAATATATTTGGAACATTAACA
71385
71425 ATTAAGGATTAACCGCTAAACACAGAATCATCTTTTCATATGAACGACATATTTT

71471
71485 TCAAACTGATCACTCATCAAAAGCATCCAAATCCGTAATTCATCACTAGCAGGCCAATTA
71471
71545 AAGCGATTTGCTTGGCCCATATCTTTCACAAAATACACAAAATCAGTCACCTCTTCA
71546
71605 AGATTTAAGCGCTTGTGCTGCCCAAGCGTATTGTGCTAAGTTTCAAGCTCTTTTATG
71612
71665 GCTTCTTTATAACGCACACTATCGCTTCCAAAGCTTGATGATTAAGGTTCAACACCTATCG
71685
71725 CACATGCTTCTAAAAAATTTCTATATCGTCAATTTTCATAATCTAATCTAATCT
71760
71785 AATTTGCTTAAATTCCTCATTTTAAACCAT
...
72182 TATTTAAACCAATCTTAGCGAAGGGGATATACCTTTTATCGATGAAATTCATCGCTTAAG
72232
72242 TCCTGCTATCGAAGAGTGTCTTACCTGCAATGGAGGATTACCGC
...
72353 TCCTATAGGGGCTACAAAGCGTGCAGTATGCTTAGCAATCCTTTGGCTGATCGCTTTGG
72403
72413 TATGCAATTTAGATTAGAAATTTTACAAAGATAGCGAACTTGCCCTC
...
72469 AAGCAGCTTTAAAACTTAATAAACTTGCGAAGAAAAAGCGCACCTTGAGATCGCTAAAA
72519
72529 GAAGTCGTTCAACCCCTAGAACTAGCTCTAAGCGCTTTTAAAAAGAGTTAGAGATTTGCC
72532
...
73470 AGCGAGGTGGGTAATGTTATGGCTGTGTACTTTACTCTATGGTATGTAGCATAATTT
73520
73530 CAAGGTGCGCTTTTGGTTTAAATCACAATATTTTATGTTATGATG
...
74061 CTTTCTTTTAAAGAAATACCTTTTATGGCATTTTGGCTACTTTAGCAGCGCTTTTCA
74111
74121 TAACCTATGTGTGATTTAGAGCGGTAACTAGCATTTAAGAAATAT
...
74673 GTTCCCACTCGGTTCTCTCTATGGCAAGTCTCTTTAAAGTTGGTAAGAGCGATGATT
74723
74733 TGCTCTTTTGAATGATAAAGCATAGAAAGATATCCACTAAATTTCT
...

...
63268 ATTTTTCGTATATGAAATTTTATAGTCCCTTAAATTTAGGCAAGATTTTGGCGATAAAT
63328 TTAGGGATAAATTATCAAAAAATTTCAAAACAATTGATGATTATCA
...
63440 TAAATATCACTATATGAGCTATTTAAACTGCAAAGCATAGTATTTTGTAAAGCCCTTGTA
63500 TACTATTAGTATTTTGGCTTTGCTTGGAGGGTGTTTTGTCTC
...
63697 TTAGCCTTGATTTTCATAAAGGCAAAATCACTTCTATAAATAGGAGCCAAATGGCGCTGAA
63757 AATCCACTCTTTTAGCTCTAGCAAGTCGTTTGATATAAACCTAGTAGTGGAGAAATTTATA
63817 TCGATGGAATGAATTTAAAA
...
63874 TTCTCAAAACAACAAAATATCAATTTAAGATTAAAGTTGAAGAATTAGTCGCTTTTG
63934 GCGCTTTTCCGCATTCTCAAGGAGACTTATACAAAGCATATAAATCAAAATCAATGAAG
63994 CTTTAGAATACATGGAACCTTAGCAACTTAAG
...
64036 TAGATACATTAAGCGGGGACAAAGCAAGAGCTTTTATAGCCATGATCATCGCTCAGG
64096 ATACTGAATTTATAATGTTTGTATGAACCTTTGAACAATCTTGATAT
...
64501 CTACAACAAAACCAACACTGCTACGGTGAAAGTTTGCCTATTAGTATGAGCGATGAGG
64561 GTGATAGTTTTTTAGTGAAGGATAGCTAGGAGAAATAAAATCCC
...
64633 TAGATCTTGGAAATTTTAGATACTTTTGTATGCTTTTAAATTAATGATAAAGTCGTTGGCG
- |
64693 TTCTCGTAAAAATTTTACCAAAATACCTACAACAAATTTTAAACAAACCTAG
64694
...

64766 TGATTTTGAAGCCATTAATGCTTTTAAACCTGATCTTATCATCATTTCTGGACGCCAAAAG
64826 TAAATTTTATGACAAATTTAAAGAAATAGCTCCAACTTTATTTGTAGTCTTGATAATGC
64886 AAATTTTAAAGCTCTTTTGAACCAATGTCTTAAAGCGTTGCAAACTTTATGTTGTTTGA
64946 AAAAGAAGCTTTAGAAAAAATTTTCAGATAT
...
65218 TCATCTTAGGCAACAAAGAACGTGCTCAAGCATACTCGATAATGCATCTTGTGCGTAATAA
65278 CCAAGCAGCGCAAAACAAAAATCATCTATCTTGTATCCAGAATA
...
65335 CAAGTGGAAATGGACTAGAGTCTTTTAAAACTATGATTTTAAATAATCAAAACGCTGTAA
65395 AATAATATACTTCGCAAGGGCTAGTCTCTTGTCTTCGATACTTAAT
...
66369 TAAAAGTCTTTTGGAGGATTCGCATTAAACATCTTTTTCATCTGCATAGTAACGCTCTCC
66429 AGCTTATGATTATTAACTACTCTTTTATAATAAGCTCTCCTGGA
...
66685 TCTAAAACCTTCTTTTAGTCTCTCGCTAATTTTTTTTCGCATCAGCTTTTCCATCTGCTTTGG
66745 AATTTTCCATATTGCGCTAGTTCATCGGTAATTTTTTTAGTATCCATAATCAAACTACA
66805 ATTGCATACTCTGCAGTTCTTTTGATCATGTGCAACGCTGTTTTTGTATGTCAGCACTTGA
66865 GCTTTTAAATAAATCTTCACITTTGTTAT
...
67945 TTTAGTTGCTCTGGAACCACTTACCCATTTTCAAAATTTGGATCATCATCGCTAAGC
68005 AAAGAAAAATCAGCATTTTGTCTGCAATGGCATTTTTTATTTGTCTTCAT
...
68062 TGATTGATATCATATTATTAGCCATAAAACCTCCCGCTATAAGTAAAAACCAACGACAATA
68122 CCTAAGCGTAAATATTTTTTTTTCATTTCATCTTCTCCTTATTAAT
...
68517 TGCAGTTATTACAAGATTTCACTTGCATTTGCTTAAATAAGATGTGCCCTTAGCGCTTATAA
68577 AAGTATAAAACCCCAACCGCAAGAAAAACAAAAAGCAAAACAAAGAA

56121 TCGCTTGTATCAATACGCTTGAATTTCCCCCAACGCTTAGGATTTTGCATCGCATCGATA
|-----|
56152

56181 TAAGTGGGTTTATAAAAAACCCCAACCA
...

56291 TGGCTAATGCTTGCATAATTTACGGTTTAAATCCCCCAATTAATCCACGAGAAACGCTTGAT
|-----|
56341

56351 AATGCTATCAAAAGCTAAATTTGGAGCCACTATAAATCTTACCCCTTA
...

56400 ATATTATAATCACTCAAAAGTAAACTTGGCGCATGAAATCCACGAGCTCATGCGCCTTGA
|-----|
56450

56460 ATTCCACCAAAATTAACACTGTATTATTATCATCTAGACTTTTC
...

57502 TTTATGGAACAAATTTTGGGTTTAAACCTTGGGATTTGCTCATTTAAAGAGCGATGTTA
|-----|
57552

57562 TCCCCACAAGCTAAAGCAGAAATGGGAAATGTTTTTAAAGCAGCAA
...

57754 CTGCTTCTCTTTTAAATAAATCTCAAGAAATCTCCAATCACTGCAAAAATCGCTACAA
|-----|
57804

57814 AAAACGAACAAAGCAAAAGCCAAAACTATAAACAANAACCTCC
...

57945 GTTTTACCCTAAGTTTTCGGATAAAATACGCCCCACTATCACAAAGCTGCAAGGATAATG
|-----|
57995

58005 ATAAGCCAAAAGTGCAAACATACCCGTGATCTAAATACACTTGCC
...

58080 ATATAAATAAGAGCTGGTTTTAAGGAAGCCTTTTTATAACCAATATCCACGACTAAT
|-----|
58130

58140 AACAAAACCTAAATCCCAAAAGCAAAGCTTTATGCGATAAACTTC
...

58233 GCAAAAAGTTTTTTTGTCTTCACTAAAGATAGATACAATAAAACCGCAAAACGATAAAA
|-----|
58283

58293 TTGATAAAAAATTTGATCAATCAACGCGATAATAATAATAGCCCTTA
...

58354 AAATAATCTAGTTGTATTAAACATATTTTCTTATATACTAATATCTAAGCGATGAAT
|-----|
58404

58414 GCTTTGCGCTCTTTAAATTTGGATCTTTTAGACTCATCTCTTTCA
...

58723 CATTTTAATCCTTTGTGATATTAATACTTTTAAAAATTTGTATAATTTACAAACGCCTTT
|-----|
58773

58783 GCTGAACCTTAAACAAAATTTTAAAGTATAATCGACCCCAATACGA
...

59202 AAATTTTACGCTTTTGTCTTAATTTTACCATTTTATAATAATAAATAGCGCTTTGC
|-----|
59252

59262 TTAGGATTAAATTCAGCTTAAATTCAGCTCTTTGCGCTTCAAAAT
...

60702 TTTTATATGGGCTAGTATTTACACAAGCTTTAGAGTATGAAAAACAAGCGCTATATA
|-----|
60752

60762 AAAAGCTATGGAAATTTACAAAAAACTCGCTTTAAAAAATTCATC
...

61461 ATTTTATCAATCTTATGCTAAGAAATAATCTTAAATTCACAATAACAAGCGCTATAC
|-----|
61511

61521 AAGTTGATCTAGGATATGATATTTTAAACGGAATTTATTTGGTA
...

61860 AACTCCACTCACTAGAAATCAATCACTCTCACAGAAATTCCTAGACTTATAGGATTTA
|-----|
61910

61920 CTCACAGGAATGAGTCTTTAGTATATGTGGCTTATCATGCAACAAC
...

62007 ATGGATTGTCTAAATTTGGCATACTTAATTTCTTTAATAATTTTACTGGAGCGCTTTT
|-----|
62057

62067 TTCACITCAAGCTATCATTGGCTTCTATATTTGCACCTTTTGGGTCTT
...

62419 TATTTTTAGGCTTAATGATTGTAAGCAATTAATCACTAGCTTAGTGTAGGCTTGGGA
|-----|
62469

62479 TTATCCCTTTTATAGGCTAAATTAATCCCAAAATTTAGTAGCTCTTA
...

62754 GATTTTATATTTGTGGGTTTAAATGGCTTTGATGAGTATGCGCTTAAAAAGCTCGTTT
|-----|
62804

62814 ACAATTTGCGCTATTATTATAGTAGTATTTGCTATGCTGCTTTTCAACAGTATTTTCA
|-----|
62848

62874 AACTTTATGCAATAACAANAATCTAACCC
...

62942 AAAGTCTCTTATCTTTTCTTTTGGTGGCTAATTTAAGTGTATTATAAAGATATCA
|-----|
62992

63002 ATTTTAAATCACTCTAGTTGTGATGGTGTCTTTTTCACCTAGCACT

47190 ATTTGATCAAGATTTGTTATAGCTGTTTCAGCTATATCCATCAGAGCATACGCGCTTTA
|-----|
47240

47250 AGTGTGTAACACCTGCTGTTTCATCTTTTACTCCAAAAGCAGTTG

...

47349 GCAGAAACATTATATACCGTAGAAAGTTGCCGAAGCAGCTGAATAAGCTATAGCGTTTGA
|-----|
47399

47409 AAACCTGTGGAATAATTTTACCCTACCTACACAGAATACCTGAC

...

47505 TAACCACCTAACACAACTCCTTTGTTGCAGAAACCAATCCCATAGCATCAGCGATATTA
|-----|
47555

47565 GCATCAATTTGTCCTTTTGACTCTCTTAAAGAAACAGAGCTTGAG

...

48487 TAGATATAGCTTGACCTAAGTATTAGCTTGAGATCTTAAACTATCTGCTATCGCCATCC
|-----|
48537

48547 CTGAAGCATCATCTGCTGCGAGTTGATTCTAAGACCTGAACCTAAGTCTGCTTAAAGAAG

48607 CATCTAAACTTTTACTATTAACTAAGCGTTTGCTTTTGCAATTTAAAGCTGCAACATTGG
|-----|
48632

48667 TGTTAATACGAAATCCCATTT

...

49041 ATAAATCTTTACTGAGTATATTTTCTAAAATTTGTTTAAAGCTAAAGCATCGTCTAAA
|-----|
49091

49101 AATCTTTAGCATATCTTGATCTTTTGTATTTTCTTTTGAATT

...

49553 ATGACTTACATGAGTGCCTGTATCGATATAGCCAAAGTCATTAAAGTTGAAGCGTTGATA
|-----|
49603

49613 AAGGCTTTATTTCTTAAACAATCATACAATTTTCAGCATTTAAA

...

50870 ATAAATCTTTACTGAGTATATTTTCTAAAATTTGTTTAAAGCTAAAGCATCGTCTAAA
|-----|
50920

50930 AATCTTTAGCATATCTTGATCTTTTGTATTTTCTTTTGAATT

...

51382 ATGACTTACATGAGTGCCTGTATCGATATAGCCAAAGTCATTAAAGTTGAAGCGTTGATA
|-----|
51432

51442 TAAAGGTCATCTCTTAAACTACACTTTTATTATCTAAAAAATGA

...

52803 CTTCTCTAAAGATTTTGAATTTTCTCTTGCGGGAGAAAAAAGCGCGCATGAG
|-----|

52863 AATTACTCGTCCAAAAAGTTTGTGCGAACAAATGTTTAAATG
52853

...

53380 TAAAATAAGCTTGTCTATATATCATCCATAAAAAAGCTTGATCCGGTTTGAGGACGATGGC
|-----|
53430

53440 AATACAATGATAAATTTACTCCAACTACAGCAGTATTAAATTCCTT

...

53857 AGTATAAAATCAGCTTTATTTATAAAAAACAAGTTTAAAGCATCTTTATCATCGCTACG
|-----|
53907

53917 CGAAAATTTACTTCTTTTATACCGTTAAATTTACTAAGATCAACATTTTCAACCCCGCTT

53977 ACATTACTCATCGTTTATATAGAGCTTAAATGATCATTTTTTAAAGTATAAATGCTTATA
|-----|
53986

54037 TCATT

...

54091 TTACTTGAATCATTTTGGCAAGCAATAGCGGCTAAATTTCCCAAAATACGATCACA
|-----|
54141

54151 AAAACAAGTCTCTAATATATAGTAAAGCTTTTTCATCTATATCCTTG

...

54352 AAATTTGCAAAACGCGAACGCAAGTTAAGATTTGCACCTAGCTCCACCTACACGCAAAA
|-----|
54402

54412 TTTTAAATTTAAGATTAATAATTTTCTTAATTTATCCATGA

...

55221 AATTTATCTTTTGATCTTGCCATTTAAGAGTATTAAATCCAAATTTTATAGCGTTTGT
|-----|
55271

55281 ATCAACTTGCCTATTGATATTGCTTAAAGTTGTTGCAAAATTTG

...

55521 GGAACGTGGTTTCATTGCCATCAACTTGAAAAGTTACTTCAATTAAGCATCGCTAAT
|-----|
55571

55581 TCTAAATCTCAGGATATAATCTTGTGTTGAAAAGGATTTGCCATATTGGATATAAAAGAA
|-----|
55612

55641 TTTCTCTCTCGGATAACAGCCGTAAGT

...

55941 GCTTTGTGATTTGAAATCACAAAATAAGTTAAAGCTTTAAGCATTTCCGTAATCCGCAAGC
|-----|
55991

56001 TCTTGGGTTTGGTGTGTTTAAAGATGTAAGATGCTTTGGATCAGCAAGCGAAGATTG

56061 ACATATTACGGTTTGAGTGGCGATATTTTCAAGATCAGGGTATTTCCTTGCATTGATT
|-----|
56081

36900 ATCGATGCTTTTGTAACTCAAGCTATCCTTTTGGCAAGATTTGGGTAACGCGCTTAT 36950
36960 TATGAACTCAATACGAACAAATTTCGGAAGCTTTAAATTTACATT
...
37312 AAGGTATGCTAGATCGAGATATATCTGGAACITTTGGTATTTTACTTTAGCGATGAGA 37362
|-----|
37372 GTAAATTTATCACAGGGCAAACTTTAGTGGTAGATGATGGATGGGG
...
38487 GTGTAGCTAATATGAATTTTACTTGTCTATACATTTAAAGCATAAATAATATCGTTTAA 38537
|-----|
38547 TAGGACAAGATTTAGCTTATGCTAAAGATGGACTTTCTCACACTAA
...
39875 GTACTCAAATACAAAATATAAAATACTCAACTGATAGAAATGATACGATTTATCGTTTAA 39925
|-----|
39935 ACAAGGTAAATGATCCACATGATCTTTAGTTGGAATCAAAACATAC
...
40199 GCATGCTAGAGCGTCATGATATAGTTCTCTAAATGTTTGTGATAATGATTTTGGCGATTTTA 40249
|-----|
40259 ACAAGGCATACTTTTTCATACTTCTCTGTGGTTCACAAAGAAGTATTAGATTTTITAG
...
40319 AAAAAATCAAAGACTTATATGCTTGATACATCGCCCTTTAAATTTTGCAGCTAGTTTAA 40349
|-----|
40379 AACTTGATGAATATGGATATCTTGA
...
41159 TTGATACTCAAACAAACCATAGAAAAAGCTTTAGAAAAATTGGAATTTTGACGATATAA 41209
|-----|
41219 TCAAAAAATAATTAGGATTTCCAATGACCTTCACCCCTTACCCAAAA
...
42167 GAATTTGATAAGGATATTGTGTTATTTGTGCTGGTGTGTTTCATCCAAAGCGGATTCAG 42217
|-----|
42227 TATTTAAGGGTAGAAATTTAGTCATAACTCAAAAAGTTCTTGCTT
...
43524 TATCTTTTGGCCTTAATTTTCATAGATTTTAGCAAGATTTAAGCTTAGGGCGACATCAT 43574
|-----|
43584 TATTTTACATCATGATGATTAATATCTTCCTAAAAATGGATAAAGTTTTTCGTTTTGATAGG
...
43644 GGATTTGTTTATCGTCTTATATATCTACACATTCGTAATGATTTTATGAACACTATAA 43655
|-----|
43704 AGAAGAT

...
44162 AATACATTTTAAATCATAAAAAGCCAGCCCTTTTGGAGATGCTTGGGGTAGACGATTTATG 44212
|-----|
44222 GATATATTACCGCAGGCATGAGTGTATCTCATATGGCTTATGAACCT
...
44495 TTAGCAAAAGTAAAACTTTATAATTCCACTGAAGGAGGTGCTAGGATAGAAAGCGCTATAG 44545
|-----|
44555 AAAAAACCTTTTAAAGAACTTTGCGAAGATTTACTCAAAAAATAAAAA
...
44839 TTTTTCGAAGAAATTTTAGGCCCTTACTTTACATCATGAACAAAGTATATTTAACGCCTTTA 44889
|-----|
44899 TATCTTAAAGATATCAAAAGATGAATCTGATAAACAATAAATAACTTT
...
44972 TGATAGAAATATCATAGAACTTTTAGAAGTTCAAGATAAAAGACTTTAAATTCGCTATTT 45022
|-----|
45032 TACCTTTGCTAGACTTCTTAGAAAAAGAAAGCACCTTTAAAAAGCT
...
45326 ATTTGATCAAGATTTGTTGTGGCTGTTTACGTATATCCATCAGCCATAGCCGCTTTTA 45376
|-----|
45386 AGTGTGGTAACACCTGCTGTTTCATCTTTTACTCCAAAAGCAGTTG
...
45485 GCAGAAACATTTATATACCGTAGAAAGTTTGGGAAGCAGCTGAATAGCTATAGCTTTTGA 45535
|-----|
45545 AAACCTGTGGAAATAATTTTACCCTACCTACAGAAATAACCTGAAC
...
45641 TAACCACCTAACACAACCTCCTTTGTTTGCAGAACCAATCCCATAGCATCAGCATATTA 45691
|-----|
45701 GCATCAAAATCTTCCTTTTGTACTCTCTTAAAGAAACAGAAAGCTTGAG
...
46623 TGTTGATAGCTTGACCTAAAGTGGCTGCTTGTGAACGCAAGAAATCTGCTATCGCCATCC 46673
|-----|
46683 CTGAAGCATCATCTGCTCGGAGTTGATTCTTAAGACCTGAACTAAG
...
47064 GCACTCTCGGCTGCAAAAGTCTACATCAGGGATTTTGGGATTTCTGCTGCTTTAACGTTTACT 47114
|-----|
47124 TGAGTTACGGTGATGTTGTTTATAGTTGATGATTAACCTTGATTTTGTGA
...

24394 TAAACAAGGAGATTTCCAATGACCTTCAACCCTACCCAAAAAGAA
...
25340 GAATTTGATAAGGATATTGTGTTTATTTGTGCTGGTGTGTTTCATCCAAAGCGGATTGAG
|-----|
25390
25400 TATTTAAAGGTAGAAATTTAGTCATAACTCAAAAGTTCTTCGTT
...
26543 GTTTTTTTTTCGGAATAATCACCAGAAATATAGATCGATCTTTTCATCTTTGGCGCTTTAA
|-----|
26593 26602
26603 -----|
TCGTATCCCTACTCCTACACAGCCCCACAAAGCTATGTGATACCAATGTCAA
26607
...
26759 TCGATGAAAACACCCCTTTACAACTCAAGCCCTTATTCTGCAAGCAAAATCGCTGCAG
|-----|
26809
26819 ATATGATGGCGCTGAGTTATTATAATTCCTTTTAAATTAATGTCAATATTGACAGCCCTT
|-----|
26826
26879 TTA
...
27324 TGTTTAAAGAAATTTCTTTTATAAAGTCTTTTATAAAGAAATATCGCCCTAC
|-----|
27374
27384 ATGAGCCTTGCTTTATAGCAATGAGAAAAATATCTCTTAGAATG
...
27468 TAACGCGTTTGAAGAAGCTTTAAAGAAAAAACCAAGCACGCTTTGTCAATCGCTACAA
|-----|
27518
27528 ATACAGGCACAGCGCCCTTCATATCGCTTTACTAGCTAATGGTATCGATGAAAAGTGG
|-----|
27551
27588 AGTCATCACTCAAGCATAAGCTTTGTGCTACGGCCCAATGCTATCGCTTATACAGGTG
|-----|
27614 27632
27648 CAAAACCGGTTTTTTTAGATATAGATGAAAACACTCTAAG
...
27762 CCACACAAACTCATCAAGCTTGTGTTATTATGCACACCTTTGGTTTAAAGCGCTCATA
|-----|
27812
27822 TAAAGCCATCAAGAGCTTTGTGAAAAATACCACATCCTTTTAAT
...
27935 GGAATTTTAAGCTTTAATGGCAATAAATCATCACAGGTGGATGCGGTGGCGGATATTA
|-----|
27985
27995 |-----|
AGCGATGATGAAAAATTTAGCCAACTTGAAGACACCTTAGCACCAAGCTAAAT

27995
...
28375 TTTTCAAACTGTCAAGCAATGAGCTTATGAACACTAAAAATTTAGAAAACGCTTGAT
|-----|
28425
28435 TAATTTACCTAGTAGCGTTAGAAATAGCAAAATAAAAGGAATGATGCGTTGCAAAATTAA
|-----|
28448
28495 TTCAAAGCA
...
29320 TTGAAAAACAAATAAAACCTTCTACAAAGAAAAATAGATTTTCTAATTTTAAAGCCGGAA
|-----|
29370
29380 AAATAAACATTTACCAACATATAGAAATCTTTAAAAATAGCTTGAC
...
29542 TTTCAATACTATTTTATATCAAAATTAATAGAACTTTCTGGTTCTTAATTTAGAACGATTAC
|-----|
29592
29602 CCCATTGAAAGATCTTTAAAAAATTAGGATTTTCATATCATTTGCACT
...
30248 TTCAGTGCATATCATGAAAGAAAAAACTTGGAAATGAATCCTATTCTTTTAGCGTTGAAG
|-----|
30298
30308 ATAATTTACCATGACTGAAGCAGGCAAGAAAAATCTTAAAAATTT
...
30530 GAGAAAATGTTAGTTATGAATATGGAGTAGTAGTACCGAAGAAACTCCTAGCGCTAAAG
|-----|
30580
30590 AAATATTTTAAATGGTGTAGCTAGTCATTTAAATATAAATGAATT
...
32093 TCGATCTTAAAGCTCATAAAGAGCTTTATCTTACACGCAAAAAATGCAATATCGCCTTC
|-----|
32143
32153 TTCTACTCCTTTTGTGATCTTGAAAGTGTAGATCTTTTAAATGAGTT
...
32300 TGGCAAAATTTAGCGGAGATAGAAGAGCTTTAAATGTGCTTTGCAAAAAATGGCGCTAAAC
|-----|
32350 32357
32360 --|
GCCAAAAATATCACTCTTTTACACTGCACACAGATACCCCTGCCCCCTTTTAATGAAGTCA
32362
32420 ATCTTAAAGCCATGCAAGCTTTAAAGAGCGCTTTTAAAGCTTGATGTGGGTTATCTGATC
|-----|
32446
32480 ACACAGGGGTATACACATTTT
...
|-----|

...
18026 TCCATAAGACCCCCCATAAATTTTGGTAAATCTGCAAAATTTGGAAAGCGTTTAAAC
18076
18086 TAAATCTTTTAAAGCATTGATATGATCAATAGCTGCCATAATGCCA
...
18798 GCATAGAAAATGCTTTAAATTCAGGTATTTTGTGATGAAGTGGTCTTTCAAGCGCATGATG
18848
18858 AAGAAATCATAGAAAGTAGCCTTAAATACGGTGCAAAAGCTCCTTTTGTGGGTGATAAAA
18918 ATTTAAGCGATGATTATGCTTCAAGCACCTGCTGTAGTGCAAAATGCCATAGAAATTTTAC
18923
18978 A
...
19141 TCTAAATGAAAAACAATCAAGTTTATATGTTTGTATGATGAAACATTTATAAAGTCGTTTCA
19191
19201 AGATCTAACAAAGCTTATCATGATGACGAGCTTTTATTTTGGT
...
19351 GGAATTTGCAAAAATACTTTATAAGGTAATCATGAAAGTGCTTTTATAGAGCGCATAGCT
19401
19411 CCAGTCAATAGGCTTTTGACACATTAACCGCATCTAGCTTTAGCTTAAACAATATAGCG
19467
19471 ATGTAAGTTTGTGCTTGGCTTTAGAGGCTCCTTTGATCGATGAAATTC
19472
...
20694 TTAGGGGTATATTTAGAAAAATGAAAAATATAAGCACCCGCTAAAGCATCGGGCCCTAAA
20744
20754 TTTAAAGCCTCAACTCCATCCACTTGGCTTTAGAGGCTCCTTTGATCGATGAAATTC
...
20984 ATCTCCTTTTCTTTTAACTCTATAGAACAACTATACATTGAGAACCAAAAGCATTACC
21034
21044 CGCTTTACTTATAAAATCCATATCTTCTTAAGGCCTTAGAATTGATA
...
21465 ATATGATCTTTTGTAAAGATGCAACAAATTTATGTCGGTATTGTGCTTTAGCGCTTACA
21515
21525 AACTCATCTTTGCAATTTACATAAAAGAAATGACAAAAATAAAAAAT
...
21787 AAATTGCTCTTTTAAAGCTTCTATAAAACCAAGTCTTTTAAATTTTTCATCGCTTCTT

21847 TAAAGATCCCACTCCAGGTAAAAACAATTTATCTGCTTTTGTCAA 21837
...
21934 TTTCAAAGCCTTAGCAACAGAGTTTAAATTTCCCGCCTTGTAAATCAATGAGCGCTATCA
21984
21994 TTTTAAACACACTCGTCTTTTCTTACCAAAGAACCATCATATACT
...
22415 AAATAATACCTTAAATAAGCCCGTAACCCCTACTCTTTGAAGTTCTTTCATCGCTTTGGA
22465
22475 TAAGTATATAAAACAATCTTTTTCAGTGATACCATTTACCCCAA
...
22993 TTCTTTCTCTCTTTCTTTGCCAATTAATATTTTGTATTTTAGCTTTGAGACGACAAGC
23043
23053 ATCACAACACCTTCTTTCATCAAAAAATGTAATCAAGGCTTAGTATCT
...
23109 ATTTTTCAAAAATTTTCATAAATTCCTTTAATTAGTCTTAAATATATATATCGTTTATA
23159
23169 TTAATTTTATTTTATGATTTTATATATTTTATATATATAAAAAAT
...
23506 CTGATGGGATAAAAAATCCAAACTTACACTCCAGATAGCCTCACACTAANTAGCGATAAAG
23556
23566 AAGATTTTATCATCAAGAGGAGACTTTTGGGACAAAAAGAAACTTTA
...
23684 CATACTTTGTGTTTTTCAAGTCCCTTTTGCTAAAGAAGATGTAGAATTTCTTAAAGCTTTGA
23734
23744 TCCTATAGCCTATAAAATCGCTTCTTTTGAAGCCAAATGATGAAAAATTCGTGGCTTTAAT
23760
23804 CGCCAAAGAAAAAAACCTACTATCGTTTCAACGGGCAATTGCTACAGAAGAAGAACTTTT
23807
23864 TAAAACTCTGTGAAATTTT
...
24136 ATGAATTTAAAGCTATGGTAGATGCAGTAAGACAAGCTGAAAGTGTGCTTTAGCGCATGGTA
24186
24196 AGTTAGATCTAGATGAAAAGGTCTTTAAAAACCGCTGTTTTTGTGCAAG
...
24334 AATTTTATCAAGAACTTTTAGGAAAAAAGCCAGTAAGATATAAATTTTGGCGATGCTT
24384

[illegible]

3810 CTACCCCTTTAAACCTTTGCCGCTACAGCTAAATCGCGCCTTAATGCTGGAGCTAAGGTAG
|-----|
3870 AGTTTATCGACATTTAAACGATGGCAATATTGATGAAAAAACTAGAAGCAAGACTTT
3887 3892
3930 TAAAAGAGAGTGAATAATAGAGCTATTAGCGTGTAGATTTCGGGGCRAATAGCCTTG
|-----|
3983 3988
3990 AAATGGATGAAATTTCAAACCTTACTAAAAATACAATATCCCTTTAAT
4191 AAATCAAGCTTTTAAGAAGCCATGTTATGTTAAAAAAGGCTTTGGGATAGCGGATATGG
|-----|
4241 4246
4251 TAGAACTTGGCTATAATTACCGCTTAAGTGATGTTGCTTGTGCTTT
5072 CATAGAGAAATGCCGATGCTATTCTTATGTAACGAGTTATTATAAAGCGCTTGGG
|-----|
5122 5127
5132 ATTTTGTCTCAGGCACACGAAGAAAAAACTCAAAAAAGGCAAA
5446 TAGAGTGTTTAAAAAACATGTTAAAGCAGCTTTGTGCTTTCTTGTAGCGGATGATC
|-----|
5496 5501
5506 ATACTTACTCACTACACACAGTTAAAGAGACACTCTAAAGGATAAAGTCGCCCTGC
|-----|
5547 5552
5566 ATACGCTCAAAAAATAAGAAAAATTTCAAAGCTTTTAGTTTTTTAGA
5749 TAATGGGTGGACTTCAAAGCATGCAAGAAATGATTTTAAATCTAGAAATCAACGCCGTAT
|-----|
5799 5804
5809 ATGAAAACACCATAGTATGCGAACCCCAATCTTGGCAAAAGAGGATT
6299 GTGCTGAATGAATTTTTTAGAAGTACGCGGGGGGGAGGTTAGCGGAAGTACGCCCTTA
|-----|
6349 6354
6359 TTTTTTCCTTGTGTGAAGGGTGTAAAGCGGATATTTTTTAAA
6710 GAGATTTTCTAGAAAGCTATGATGCTAAAGAATATGTCAAAGCCATAGGCAACGCCCTTA
|-----|
6760 6765
6770 AGCAGGATAAATTTGCCATATATGAAAAGCTTCTTTTATGAAAAAC
6865 ACTCCACTTAGCACTTCATAGAAACTTTTACAACTAGGACTACAAAAGGCGATACCT
|-----|
6915 6920
6925 TGTGTGTCATGAACTTTTAAATTCGGTTTTCCCTTTCCTTTC

...
7108 TGGGAGCATTAATGAATATTTTCGAACAAACAGGAGTAAACGCCACAGCATCCTA
|-----|
7158 7163
7168 TATTTTCTTTTCTATAAAGGAGCAAAAGAGAATTTATTTTAAA
...
7486 CTAAAAATTTTAAAAAAAACAATTTTTCAGGAGATCATATAAATATTACACGCTAAAG
|-----|
7536 7541
7546 AATATATAGAAATTTTATGGAAAAAATTAGATATAAACCAACAAT
...
7789 AATTGAAATCAATTTTAAACAAGAGATTTTAAGTCTCTTAAATCTCAAAAGCGCTTAC
|-----|
7839 7844
7849 TAGAAAAATTCAAACCCCTAAAGGCAGAAAAATGAAGAAAAATTTGAAAAATTTATTTCA
|-----|
7913 7918
7909 AGCCATCGCTCAAAAAAATCTTATCCATTTCAAATATCTTAAAAAATTTGAATTTTACA
...
8013 AAGATGGTATCAGTATAAAGAACAATCGGATTTGTTATTATTATGATCCTTAACGATATT
|-----|
8063 8068
8073 TAGAAGAAACAAAAATTTTTCATAGAAAAATGGGCTTATCGTTATTCTTAGCTTTGATGAAG
|-----|
8109 8114
8133 TTAAGGATAAAGTTTATTTTAATGAAAGCTAT
...
8508 AACACGTAGAAAAACCTCTTGAAATTTCTTAAAGATCTAAAGAACTTTTAGCGATGGAG
|-----|
8558 8563
8568 GAGAAATTTTGCAACTATCCCTATCAATGCTCCTGCTATTGATCA
...
8920 CTTGTAATGAAGCATTTCCACAGTATAGATTTGAACTTTTGCAAAATTTTAAACGATAAAG
|-----|
8970 8975
8980 GACCACATAGTAGCTATTGAAAAATTAATACCAAAATGTATCATTTTAGCCTATGAAAGTA
|-----|
9045 9050
9040 AAAATATCGCCCAAGATTTAATTTCTTTTGAACAACAAGGAGGCAATTTGTATTGTTCTTA
9045 9050
9100 T
...
9978 CAAAAGTGAAATATCTTAAAAATTCCTAAAAAATTTAATATAAATACTTAACGCTATGC
|-----|
10028
10038 TTTTATAGATGATAATATCGCGAGCTTGAAAAATACCAAAATTTACAGCGTTTAAACAC
|-----|
10055
10098 TTTTATCGGATGAAATATACTTTTATAAAATCAAACTTTTTC

```
to HP0462"
/codon_start=1
/transl_table=11
/product="putative type I restriction enzyme S protein"
/protein_id="CAB73967.1"
/db_xref="GI:6968970"
/translation="MNNWKKKLGIDIAEITWGQSPKSEFYNFONIGMPFLQGNRTGR
KYPFDTCYEYKKAKKGEILFSVRAPVDGINFANNDICIGRLGSMNAKNGENEF
YVLLHNRSLVINNESGVFSVKNKDLOTIEILLPLEEOROIATILSSIDDKIEL
HEQNKLELAQTLFLNWFKDRFNSTISDFISMQNGFAFKSKDFIDYGNNGVIKKN
ISGIVDIVNTKISONTINEVNKNFNINSGDILFAMTGAIEIKMGKIVPSTNKKLWN
ORVGMKRELFGARFLAYIHLTSEFGYDVYINSATGSAQENISATDIENGPFVKLTSE
EIVYSKQLNDFEKEKIFNLGEIQTLNMRDLILPKLLNGEIKITN"
BASE COUNT      91877 a 43965 c 35199 g 92294 t
ORIGIN
```

CJ11168X5 Length: 263335 February 21, 2002 16:59 Type: N Check: 8802 ..
Found using 'seq2-3' (pappu403.key)

```
...
70  TATTATTAAATAATCAGGATTTTCCATCATGGAAGATGAAATCTTTAGTCGTTTTTA
    |-----|
    120 125
130  TCCCACTTATAAGCAACTCATCTAAAGCTACTTTTCATCTTTTGCAAT
...
614  TGTTTTTTTCATCAAGCAAAATCGCAGGGGATCTCTCTATAAGTTTTTGATGGCGTCTTTG
    |-----|
    664 669
674  CATAGACAATCTCTTCGCCACATGAATCACATTCGCCAAAACATA
...
735  ACTTCTATATGGCGTGGATTTTGATGTATTTTCCATATACATAGTTCGATCGCCTTCTTA
    |-----|
    785 790
795  GCTGTCATAGCCTCGCTTTCAGCCGACCAATACGCTTTTTCAGAT
...
912  CCTATTCTTTGGCAAGTTTTTTAGCTGCTTCAGCTCCTGCCAAGCCCATCGCTTCCT
    |-----|
    962 967
972  GGGATCACAGGTACGCTGCTCTTTGCGTACCTTGTTGCTTTTGGCTTTTG
...
1255  CTTTATCTGCTTCAGAATACACAAAATTCGTTTTTTTACCCATTCCTTTAATCGTCTCTTA
    |-----|
    1305 1310
1315  AAGCACGAGTGTATTTCTCCAGGATTAGCAATTAATATGCTTTT
...
1390  AATAAAGGCATTCCAAATTTCTACAGGCTGACCATCAGCAACTAAATTTTCAGCGATCCTA
    |-----|
    1440 1445
1450  CAATCAAATCCGCTTCAATTTTCATTCATGATTTTTCATCGCTTCGATAATAGCTATAGTA
    |-----|
    1486 1491
1510  TCGCTTTTAAACCGTACCTACCTACTTTTACAAAAGGTGCTGCACCTGCACTTG
    -----|
    1514
```

```
...
1600  ATACTTGGTTTGTGTTGATTTTACACTTTGACTAGTTTGAGCTTCGTTTACACACGCTTACA
    |-----|
    1650 1655
1660  TTGATAGTTGTGTGAGCTGGAACCTGGAGACAAACAGAGAGCTGGAA
...
2146  CTGTAGTGGATGCCAAAAAATTTCTAGAGAAGAAATGTAGTAGATTTTGAAGCGGATGTAT
    |-----|
    2196 2201
2206  GTATAGTTCTTGCAAAATCTTTTGCCCTAGCTAGAACTATAGAGTA
...
2395  CTGCAAAAATTTATGCCAATGAAGGTATAGCGCAAGTGTGTTTTTTTGCAGGCGGATGAAA
    |-----|
    2445 2450
2455  AATGTACACCACTTATAAAGATAAAAAAGGCAAGTATCAAGCCCA
...
2675  ATAAGATTATCATTATTCTCGTGATGAGCTCAAAACAATTTGAAATGGCAAGCGTTTTTA
    |-----|
    2725 2730
2735  ATGCTCCTTCTATGCGTTATTTTCATAGGAGATGTAAGAGATAAAGACGCTTTTAAGTCTG
    |-----|
    2780 2785
2795  CTATGCGTGATGTGGATTTTGTATTATACATGCTGCTGCAATG
...
2861  ACCCTATGGAATGTATAAAAACAACATACACGGTGGCGCAAAATGTCTATCGACGCTTGTT
    |-----|
    2911 2916
2921  TTGAAAATGGGTTAAAAAATGTATCGCTCTTAGTAGCGGATAAGGCTTTGTAATCCTGTAA
    |-----|
    2944 2949
2981  ATTTATACGCTGCAACCAAACTTGCAAGCGATAAGCTTTTTTGTGCTGCAACCAACATAG
    |-----|
    3007 3012
    3022 3027
3041  CAGGTAACAAGCAACACACGCTTTGGTGTTCACAAGATA
...
3323  AAATCATAGGCATAAGAGCAGAGAGAAAACTACATGAATCATGATTTCAAGGATGATA
    |-----|
    3373 3378
3383  GCCATTAAACCTATGAATTTTGAAAACTACTATGCTATTAAGCCCCGAG
...
3573  TACTGAAGGTTTTTAAATGCTTACTATTCTCATCAAAAACATCGATCAAGGATATAG
    |-----|
    3623 3628
3633  ATACGCTTACAAAAGCTTTTAAAGATGAAATTTTAAACAGGGGGTAA
...
3750  CCACCTCAGCCCTTCATCTTTCCTACACAGCACTAGGTGTCCAAGAAAAAATCGTTTAA
    |-----|
    3800 3805
```


/product="putative anion-uptake ABC-transport system
permease protein"
/protein_id="CAB73955.1"
/db_xref="GI:6968958"
/translation="MKLLILKRIFLDYFDGFKQALFLLFNADESVISAIKTTLLSS
SISIVALLIGPPLGFIIGFFEFKLRFKIKLIVDTSLSPFTVAVGLIYALISSRGLP
GEGFLTIKALILGQFIALPFIIVIALFSNLNENMKHFLLIKSPHLSPLKVLVTMI
YELRFALISVALYGRIVAEVGVAMIVGVNKKYDTRTITTTAISLETNKGFEASGIAL
ALVLIILAFCLNFITHKLKRT"
250017..250826
/gene="Cj1540"
CDS
250017..250826
/gene="Cj1540"
/note="Cj1540, probable periplasmic protein, len: 269 aa;
similar to TR:Q26579 (EMBL:AE000832) Methanobacterium
thermoautotrophicum MTH479 (293 aa), fasta scores; opt:
416 z-score: 475.8 E(): 3.5e-19, 37.0% identity in 273 aa
overlap. No Hp match. Contains probable N-terminal signal
sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73956.1"
/db_xref="GI:6968959"
/translation="MKIISIALALALASAAELKMATTTSDNTGLLDALKPLYEKE
SGNTLKWAVGTGAALKMGEDCNADVLFVHSPKAKEFMKGFGVDRTPVMVNDPIII
ADKLSAKFKGNLKELELKNELKLFISRGDSCTONKEKSLAKNLGGVPEKQSWY
QOSGGMLASIKIAEKKGVILTRGTYIKYEANEKGPALNVLVNEGDDSLKNFYSVI
ATNPKHKNVNYTEASFKIFKWTSDTLNFIADFLLNKLPLFVIDAKTRKD"
250968..251735
/gene="Cj1541"
CDS
250968..251735
/gene="Cj1541"
/note="Cj1541, unknown, len: 255 aa; similar to
hypothetical proteins e.g. YCSF_BACSU (211 aa), fasta
scores; opt: 736 z-score: 875.7 E(): 0, 59.9% identity in
177 aa overlap, and to LAMB_EMENI Aspergillus nidulans
lactam utilization protein LAMB (262 aa), fasta scores;
opt: 405 z-score: 485.4 E(): 1e-19, 44.9% identity in 127
aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1541"
/protein_id="CAB73957.1"
/db_xref="GI:6968960"
/translation="MFKVDLNSDLGSEFGAYKMGWDEILKFVSVNVACGFHAGDPC
WDETFLNAKQNGVCIGAPHSYDILLGFGRRNMQISFEAKNYALYQLGALFGFAKAK
GMKIQHKAHALYMAAIDENLALALCEAVASFDENIIFLGLSNAMNEAAKKGLR
YANEVFABRAYNDGDTLVSRKLEGALIHENLAIKRVIKMIKESKVTISINGKEIDLKA
DSICVHGDNALAEFVKIKENLKEQIQICALENFI"
251745..252485
/gene="Cj1542"
CDS
251745..252485
/gene="Cj1542"
/note="Cj1542, unknown, len: 246 aa; similar to
hypothetical proteins e.g. YCSJ_BACSU (578 aa), fasta
scores; opt: 569 z-score: 690.1 E(): 4.1e-31, 41.8%
identity in 237 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1542"
/protein_id="CAB73958.1"
/db_xref="GI:6968961"
/translation="MFSVHFGSKALLIRFQEISPOINAYVLSTEQRIQALKEGEI
YGLDELSAYASLLIYFNPCLVLSNLLDFLEKIKKDLKLAQNSLCLIEVPLCYDDEE
FGLDLEFYKKNQISKEELISLHTRPYLVFMGLPMAGFPYLGGLDERLFTPLRSSSR
AKIAGSVGIADKQTGVYPISSPGGWQIARTPLFEFFDKEDEKNPTLLKAGMFLKFKA
ISKDEFFDIOEQAVAKVYQKEIYEYKNH"
252466..253431
/gene="Cj1543"
CDS
252466..253431
/gene="Cj1543"
/note="Cj1543, unknown, len: 321 aa; similar to

hypothetical proteins e.g. YBCK_ECOLI (310 aa), fasta
scores; opt: 553 z-score: 652.4 E(): 5.1e-29, 32.7%
identity in 312 aa overlap. Also similar to part of
DURL_YEAST urea amidolyase [includes: urea carboxylase;
allophanate hydrolase] (1835 aa), fasta scores; opt: 307
z-score: 353.8 E(): 2.2e-12, 28.5% identity in 333 aa
overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1543"
/protein_id="CAB73959.1"
/db_xref="GI:6968962"
/translation="MSIKIIEASINSSLODFGRKKFAKFIARSGAMDEDLARMANIL
LGNKDEAGIELCLGKGVFLDENYFVLGSAEFAKLDNOKIKTKVYQANKGDILE
LDLAKIGRGYLCVAGGPEVKSFLNSKSDAKMGAGVEGRALQKDDILNTHNTRIFE
NLEARECNPLFKPEPIRVLIGTNDADATQKGIETFLNTTYKVLGKSDRMALYAE
SSEIHEKNSADIISDPAVFGSIQPKSGPIILMAGRQSTGGTYKIATVIENDLSLL
AQAKLGSFQFSISMQALELYKQRMKFKAMDQKINLDFENLI"
253428..254324
/gene="Cj1544"
CDS
253428..254324
/gene="Cj1544"
/note="Cj1544c, probable integral membrane protein, len:
298 aa; similar to hypothetical membrane proteins e.g.
YYAM_BACSU (305 aa), fasta scores; opt: 244 z-score: 300.4
E(): 2.1e-09, 22.3% identity in 300 aa overlap. No Hp
match. Contains Pfam match to entry PF00892 DUF6, Integral
membrane protein"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73960.1"
/db_xref="GI:6968963"
/translation="MTRKYLYLIVIAMFLWGSSWPTSKILANTDTSIITFWREFF
VYLGFFVNLFKIPDKIEKTALKWVLGVLNGLYAVFFIATKHLAGKGVLTVT
MIPIYSILFIAMIAFPQKNKSTHRTKSEILGLFLGLSGCLNLSGLDLQFKFN
ILFTCSFIWALMAVENHKAHGAHPLNFYINLISILMFSWILFPDKSYEIFHFEFK
FWLSMFVVAFLSTIIGTSIYYGHIILGSVRANSFVLITPASALICSYFILDVEPNAL
TLGCGALGAIYFINIYKKA"
254331..253823
/gene="Cj1544c"
misc_feature
/note="Pfam match to entry PF00892 DUF6, Integral membrane
protein, score 55.00, E-value 1.7e-12"
complement(253845..254279)
/gene="Cj1544c"
/note="Pfam match to entry PF00892 DUF6, Integral membrane
protein, score 2.10, E-value 0.32"
complement(254357..254935)
/gene="Cj1545c"
CDS
complement(254357..254935)
/gene="Cj1545c"
/note="Cj1545c, unknown, len: 192 aa; similar to e.g.
MDAB_ECOLI modulator of drug activity B (193 aa), fasta
scores; opt: 777 z-score: 934.0 E(): 0, 57.8% identity in
192 aa overlap. Also weak similarity to many
boxidoreductases e.g. YCAK_ECOLI putative NAD(P)H
oxidoreductase YCAK (196 aa), fasta scores; opt: 124
z-score: 159.3 E(): 0.15, 24.7% identity in 174 aa
overlap. 65.1% identity to HP0630"
/codon_start=1
/transl_table=11
/product="MdaB protein homolog"
/protein_id="CAB73961.1"
/db_xref="GI:6968964"
/translation="MKNILLINGAKEFNGSKQLNLTLLHNALETLLKTLGVEVDGTHI
DQGYDPKREIQKFIKADAVIYQMPAMWGPWIVKKYIDFVGLGAGLVFKNDGRTHE
NPSKNYGDHGHKKYMFSLTNAPLEAFNDFNDFEFGKGVDMVYVHLHKAHEFIGM
KALPTMCDNVKNOPEVKYLEVELHKKIF"
255035..255382
/gene="Cj1546"
CDS
255035..255382
/gene="Cj1546"
/note="Cj1546, unknown, len: 115 aa; similar to

```

/note="Cj1535c, pgi, possible glucose-6-phosphate
isomerase, len: 406 aa; similar to e.g. G6p1_BACSU
glucose-6-phosphate isomerase (EC 5.3.1.9) (450 aa), fats
scores; opt: 469 z-score: 569.6 E(): 2.1e-24, 29.3%
identity in 382 aa overlap. 26.1% identity to hpl166.
Contains PS00174 Phosphoglucose isomerase signature 2, and
pfam match to entry PF00342 PGI, Phosphoglucose isomerase"
/codon_start=1
/transl_table=11
/product="putative glucose-6-phosphate isomerase"
/protein_id="CAB73951.1"
/db_xref="GI:6968954"
/translacion="MLNLTFFKQSEIHTTISYANRINDEKSGDIGYHILDSLNL
IDESLQIQDKYKYNVLVGMGSSCGVRAKRLDMLEKNSQRELFVFDITSHSSEN
KTLBKILLESFLIISKTGSTIEVSLFKLRIEHLDMQELKKYFVFDTKDKSLH
QEGNLGKICFFTPANVGGFSILSAGVIVPLFCFGYNAKALLEGACAKACFEDFTHKR
DEILQKAYHYCTHKNANINVLFSYDAFKGPNENYIQLIAESLKGKGYKRIGLTPLA
LIGARDQSHQLJMDGPKNKTVTFLKIDAKAKAPIIDPIHFKFLDLSLNKNVHLELL
NAOCDATMHALLIAENLSVDVTELEKLDHAWHAGLYMYIYELFTSCGVMLGINTYDQPG
VEVCKLILKNLNS"
/complement(244033..244185)
/gene="pgi"
/note="Pfam match to entry PF00342 PGI, Phosphoglucose
isomerase, score 42.20, E-value 2.3e-10"
/complement(244054..244107)
/gene="pgi"
/note="PS00174 Phosphoglucose isomerase signature 2"
/complement(244366..245073)
/gene="pgi"
/note="Pfam match to entry PF00342 PGI, Phosphoglucose
isomerase, score 73.30, E-value 1.3e-18"
/complement(245238..246062)
/gene="galu"
/complement(245238..246062)
/gene="galu"
/EC_number="2.7.7.9"
/note="Cj1536c, galu, probable UTP--glucose-1-phosphate
uridylyltransferase, len: 274 aa; highly similar to many
e.g. GALU_PSEAE UTP--glucose-1-phosphate
uridylyltransferase (EC 2.7.7.9) (279 aa), fasta scores;
opt: 1245 z-score: 1547.3 E(): 0, 65.9% identity in 270 aa
overlap. 61.3% identity to HP0646. Contains Pfam match to
entry PF00483 NTP_transferase, Nucleotidyl transferase"
/codon_start=1
/transl_table=11
/product="UTP--glucose-1-phosphate uridylyltransferase"
/protein_id="CAB73952.1"
/db_xref="GI:6968955"
/translacion="MLQTCIFPAAGYCTRFPLPATKTLPKEMLTLLTKPLIHGVDEAL
EAGNENGVFTGRCRALEDYFDISYLEHQISGTTKKEYLLDEIRSLNCKTFFTFTRQ
NQMKGLDAVLKGRPLVGDFAFVILADDLCVNEEGNVNWAQMKYIKERYCRTIAYM
EVPKEQVSNYGVISGNEVEENLIMVNSIMEKPSDEAPSNLAIIGRIYILTPDIFGILE
NTRAKNGEITQLDALTQATNGMVLAYKTQGRFDCGSVEGFVEATNYFYKSKC"
/complement(245247..246038)
/gene="galu"
/note="Pfam match to entry PF00483 NTP_transferase,
Nucleotidyl transferase, score 32.60, E-value 2.7e-09"
/complement(246144..248117)
/gene="acs"
/complement(246144..248117)
/EC_number="6.2.1.1"
/note="Cj1537c, acs, probable acetyl-coenzyme A
synthetase, len: 657 aa; similar to many e.g. ACSA_ECOLI
acetyl-coenzyme A synthetase (EC 6.2.1.1) (652 aa), fasta
scores; opt: 2026 z-score: 2329.3 E(): 0, 48.9% identity
in 619 aa overlap. 69.3% identity to HP1045. Contains
PS00455 Putative AMP-binding domain signature, and Pfam
match to entry PF00501 AMP-binding, AMP-binding enzyme"
/codon_start=1
/transl_table=11
/product="acetyl-coenzyme A synthetase"
/protein_id="CAB73953.1"

```

```

/db_xref="GI:6968956"
/translacion="MLNQNQELFKPSKFEFSRNARIKNLCLEYDILCDKABKDFEGFKW
RQAFKEIEMFSPFARNEDKAPFYKNPEGCTLVNSVQCLDRHMKTRNKAALIFEGE
MGDEVYTVRLLHETCKAALLKFKGKGDVVIVYMPIMETAIYNLACARIGAIH
SVVFGSGPEALRDIIDAGAKLVVADGAFRCRCKPYMLKPAVDKALSEGSEVKVL
IVIRNNEPIEIKGRDYVYNELVNESKCEPEIMDSDDLFLLYTSGSTGCKPGVMH
ASAGYILWAQTMEMWDFDKIDYDNTWCSADVWITGHTYVYVYGLACGATTIMHEGTP
TYSNGRWRMIEEYQISKEYFTPTAIRMLHADAPNEPRKYDLSLTLEVLGTVPGEINP
SAWKFYDEIGGTSKPIVDTWMQTGTGMITPLPGATPLKPGCATLPGCIFAIEVID
EENKDEGEDGLLICITKPPWSMIRGTWGNDEYIESVSOAKKDGKAVYSGDGAFY
DKNGYIITGRTDDVNVAGHRIGTAEISAIKHPSPAESAVVSILDTIKGESILFAP
VVLSPASCDLGAETLKLDELNLRLVEIGKIAKELIYTPGLPKTRSGKIMRRILR
TIAGEEIKQDITLSDSKVVEIVKLAKAEFE"
/complement(246465..247790)
/gene="acs"
/note="Pfam match to entry PF00501 AMP-binding,
AMP-binding enzyme, score 371.80, E-value 6.9e-108"
/complement(247302..247337)
/gene="acs"
/note="PS00455 Putative AMP-binding domain signature"
/complement(248255..249250)
/gene="Cj1538c"
/complement(248255..249250)
/gene="Cj1538c"
/note="Cj1538c, possible anion-uptake ABC-transport system
ATP-binding protein, len: 331 aa; similar to many e.g.
LACK_AGRRD lactose transport ATP-binding protein LACK (363
aa), fasta scores; opt: 319 z-score: 386.4 E(): 3.3e-14,
28.8% identity in 208 aa overlap, and UGPC_ECOLI
SN-glycerol-3-phosphate transport ATP-binding protein (356
aa), fasta scores; opt: 298 z-score: 361.6 E(): 8.1e-13,
32.1% identity in 196 aa overlap. No Hp ortholog. Contains
PS00017 ATP/GTP-binding site motif A (P-loop), PS00211 ABC
transporters family signature, and Pfam match to entry
PF00005 ABC_tran, ABC transporters, Putative
identification as anion-uptake is due to similarities of
associated permease Cj1539c"
/codon_start=1
/transl_table=11
/product="putative anion-uptake ABC-transport system
ATP-binding protein"
/protein_id="CAB73954.1"
/db_xref="GI:6968957"
/translacion="MIEISNLFYQNKVEVLKINLKLDTSKISILMGANGSGKSTFL
RILFLFEGDSKNISYFCGNPKNKKQREIYLLPEPILLNRSVRANFLTKTYGIG
EDIERIKESLMFLNDELSSLKHPNELSSQSQKIAFAIALSVRAKYLLDEPSAFL
DKNTTLFKKTKILKWHENFTGLIASHDKHFLDSLAKKLYLHSGETLEPENTNVCE
LENGOVKFCNFIDFSCKKYDKFKPPSKIAIDPYKISFFNSKNIPKNYDFFILEKY
IIALRSKSDVFIIVSCMDKILEFALEKQBELFDLKLIELSLYFFEDATCFULN"
/complement(248633..249172)
/gene="Cj1538c"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporters, score 87.80, E-value 2.2e-22"
/complement(248819..248863)
/gene="Cj1538c"
/note="PS00211 ABC transporters family signature"
/complement(249128..249151)
/gene="Cj1538c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/complement(249252..249971)
/gene="Cj1539c"
/complement(249252..249971)
/gene="Cj1539c"
/note="Cj1539c, possible anion-uptake ABC-transport system
permease protein, len: 239 aa; similar to many involved in
anion uptake e.g. CYSF_ECOLI sulfate transport system
permease protein (277 aa), fasta scores; opt: 235 z-score:
293.7 E(): 4.9e-09, 23.6% identity in 220 aa overlap, and
MODE_ECOLI molybdenum transport system permease protein
(229 aa), fasta scores; opt: 201 z-score: 253.9 E():
8.1e-07, 25.2% identity in 214 aa overlap. No Hp ortholog"
/codon_start=1
/transl_table=11

```

phosphoribosylformylglycinamide cyclo-ligase, len: 329
aa; similar to many e.g. PUR5_ECOLI
phosphoribosylformylglycinamide cyclo-ligase (EC
6.3.3.1) (344 aa), fasta scores; opt: 1055 z-score: 1258.2
E(): 0, 52.7% identity in 330 aa overlap. No Hp match.
Contains PS00070 Aldehyde dehydrogenases cysteine active
site, and Pfam match to entry PF00586 AIRS, AIR synthase
related proteins"
/codon_start=1
/transl_table=11
/product="phosphoribosylformylglycinamide cyclo-ligase"
/protein_id="CAB73945.1"
/db_xref="GI:6968948"
/translation="MKISYEDAGVISIDGNTFFVEAIKPLVKETFNDNVVGGISFAGA
FRMPGFKPKVILGATDVGTKRLAIDAKKYDTIDGDLVAMCVNDLICNFATPLFL
DYATAKLEVEKAVVSGIAKCKMANCALIGGETAEMGMYAKDDFDLAGAVGMA
LEEDRIYRVKNGDILLALSSGLHSGYSLARKVLFSGKLFKDDKIEGKNLIDIL
LEPTRIYRDLTKPYISALAHITGGGLVENPRVLPRMGATIRKHHKLTPEIFYT
IGQAVESEMYRSPNMGVGLVMVDPNSVSKILENSDAFIIGETICINEGIVLE"
complement(239254. .239289)
/gene="purM"
/note="PS00070 Aldehyde dehydrogenases cysteine active
site"
complement(239263. .240177)
/gene="purM"
/note="Pfam match to entry PF00586 AIRS, AIR synthase
related proteins, score 289.80, E-value 3.5e-83"
240289. .240894
/gene="Cj1530"
240289. .240894
/gene="Cj1530"
/note="Cj1530, probable ATP/GTP-binding protein, len: 201
aa; similar to hypothetical proteins e.g. YTAG_BACSU (197
aa), fasta scores; opt: 346 z-score: 404.6 E(): 3.2e-15,
33.0% identity in 188 aa overlap. 37.4% identity to
HP0831. Contains PS00017 ATP/GTP-binding site motif A
(P-loop), and Pfam match to entry PF01121 UPF0038,
Uncharacterized protein family UPF0038"
/codon_start=1
/transl_table=11
/product="putative ATP/GTP-binding protein"
/protein_id="CAB73946.1"
/db_xref="GI:6968949"
/translation="MKNAPFVTSIACGKSTFTSIANSLGFKSISADKIAHKILDENA
LELEKTFSPSLNKLKKEKIDRKILGEIVFNKKAKILENTHPKIPAKTILEQMO
ILDKENAKFVEPLFEFGAYENLGVIVITPKLSLARKIMQDKLSLEAAKRLD
SQDIEKLRKADFIITKNTSYADFRQECVKVIOEISKGNM"
240313. .240336
/gene="Cj1530"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
240319. .240840
/gene="Cj1530"
/note="Pfam match to entry PF01121 UPF0038,
Uncharacterized protein family UPF0038 , score 118.60,
E-value 1.2e-31"
240891. .241640
/gene="dapF"
240891. .241640
/gene="dapF"
/EC_number="5.1.1.7"
/note="Cj1531, dapF, probable diaminopimelate epimerase,
len: 249 aa; similar to e.g. DAPF_BACSU diaminopimelate
epimerase (EC 5.1.1.7) (284 aa), fasta scores; opt: 212
z-score: 260.8 E(): 3.3e-07, 27.1% identity in 280 aa
overlap, and DAPF_ECOLI diaminopimelate epimerase (EC
5.1.1.7) (275 aa), fasta scores; opt: 195 z-score: 240.8
E(): 4.3e-06, 24.4% identity in 270 aa overlap. 41.6%
identity to HP0366"
/codon_start=1
/transl_table=11
/product="putative diaminopimelate epimerase"
/protein_id="CAB73947.1"
/db_xref="GI:6968950"

/translation="MKFYKYCASQNDFTVNADRKEDRSALAKELCNRYEGIGADGFI
VILPHERYDFEWFEYNDTEEGRAACMGNGSRAAAHFAHINKINPNMFSFLTGCAGTIKAK
VNODKSVSLGKIKSVONTFEELGKTWQLCNTGVPHLVHFHCONDEDTMLCCMKROK
YANVNVKILDENHLKRVYRGVEDETLACGTGMGACFYLAFLNKKVQNKVITPK
SGEYVGAYKNEELFEGKVYCFEANYNF"
241615. .242346
/gene="Cj1532"
241615. .242346
/gene="Cj1532"
/note="Cj1532, possible periplasmic protein, len: 243 aa;
some similarity to N-terminus of BAX_ECOLI BAX protein
(274 aa), fasta scores; opt: 240 z-score: 286.1 E():
1.3e-08, 38.3% identity in 120 aa overlap. No Hp match.
Contains possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="possible periplasmic protein"
/protein_id="CAB73948.1"
/db_xref="GI:6968951"
/translation="MKPIIIFFLPLFADDLKSGFGEYYKLDIDOKROIFFIKM
NEFDQSFKEIOERAFIEAFFKDAYKTGPTSNQINLEKITIKKYRIENLYDFAE
YKRIQIKPSMGIAQALVESATGTSRFAEANNLFGEWGEGKLLDRLHPDKHK
IKIFDSIQSDSVSYLNLNHFAYKFRDARAKFESGEGKITGLEARTLDSYSEKRG
YYINLITIKRYNLEKYDTSNNT"
complement(242343. .243380)
/gene="Cj1533c"
complement(242343. .243380)
/gene="Cj1533c"
/note="Cj1533c, unknown, len: 345 aa; 27.9% identity to
HP1321. Contains possible helix-turn-helix motif at aa
191-212 (Score 1065, +2.81 SD)"
/codon_start=1
/transl_table=11
/product="putative helix-turn-helix containing protein"
/protein_id="CAB73949.1"
/db_xref="GI:6968952"
/translation="MKVLNFEYENHPKFEYSYERKNOISKPNILIKGPRFCGKKTLPF
NFLSQKASEILFDLYDTRFEKQSLADFLNENLQILCYLDFIPNLEKINI
PIILSTNIKDLNVNGFEELDYDEFEELSVSKNLPINNLGLFLQSGRSKEGKN
ILLRQSFTELEILKYLALNGLQOISIKFTELKKRLKTSKDSVQALKKLENTYV
YITLKDEKKLQIYFKDFGLRNLCISKDFSHLFENLVISLSELFKEEFYFNKYFNF
YSQISKIAYITSSPTLDIDLKLRAKLLPALKLELGIFHVIFITLSSDNEFFEQGVKFE
VISFDFPSLGF"
complement(243445. .243894)
/gene="Cj1534c"
complement(243445. .243894)
/gene="Cj1534c"
/note="Cj1534c, possible bacterioferritin, len: 149 aa;
similar to many members of the DPS family e.g. NAPA_HELPY
neutrophil-activating protein A (-HP0243)
(bacterioferritin)(144 aa), fasta scores; opt: 397
z-score: 495.1 E(): 3e-20, 41.3% identity in 143 aa
overlap, TPFI_TREPA antigen TPFI (176 aa), fasta scores;
opt: 296 z-score: 372.0 E(): 2.1e-13, 33.3% identity in
144 aa overlap, MRGA_BACSU metalloregulation DNA-binding
stress protein (153 aa), fasta scores; opt: 234 z-score:
298.1 E(): 2.8e-09, 33.1% identity in 127 aa overlap, and
FRI_LISIN non-heme iron-containing ferritin (156 aa),
fasta scores; opt: 226 z-score: 288.3 E(): 9.8e-09, 31.8%
identity in 148 aa overlap"
/codon_start=1
/transl_table=11
/product="possible bacterioferritin"
/protein_id="CAB73950.1"
/db_xref="GI:6968953"
/translation="MSVTQLQLOMQADAHLLWKVFNHWNKGLQFSTHYTEKAY
EEMAEILFDSCAERVQLGKATICQKLVIMENAKSPKVAKDCFTPLVEYLIIKQDIYEL
LAFKKLINEAAEKESDTTTAAFAQENIAKYEKSLWMIGATLQACKM"
complement(244024. .245244)
/gene="pgl"
complement(244024. .245244)
/gene="pgl"
/EC_number="5.3.1.9"

codon 181 complement(1448602. .1448604)"
/codon_start=1
/transl_except=(pos:complement(226338. .226340),aa:OTHER)
/transl_table=11
/product="putative formate dehydrogenase large subunit
(Selenocysteine containing)"
/protein_id="CAB73932.1"
/db_xref="GI:6968936"
/translation="MSSVGENIKLTRESFLKMAALSSLATPLLARSETLREASDELK
EAYEGSKVITVCTACSVGGGIIAEVQVGVVROEIAQDPVSGGHCCKGSDMIDWV
RSHVRLKPMKENGEMWRISYEALSEIGEKLAAAYKRENPESVMFSGSAKLNNEQAY
YIRKFAAFFETNNVHQARIXHSAIVGANTFGYGAMTHLGDIOQRSCIIIIIGAMP
AVHPFVREFHLAKAKGAKLIVVDPRTKSAAKADIYARIRPGTDIAFYGMKLIIIF
DEGLDQKYLDERSVFGIDKIREAAKVTVEVENVTGISKELLVOITHEVAKNKPTEL
IWAMGLTQHTVGTSTNRLAPIVQMLNIGCKFGGVNIIIRGHDNVQASDMACLSENL
PGYPLNEATWRYAKIWTGVYDEWHLGNFVSKDWMHKTGLSLARWAAALNGDGDNA
IDNAGTEPKALVWNGSITTAQCVKVEGLEALELELLVADPFVNEAGIIAERKDGII
LLPAAOTFETSGSVTATNRSGQWRKVDPLYESMEDQEILFELAKKLGYEDFTKTL
RDEKGEIWPENATREIAKVRISIGLNGWSPERLKHHTLYWDKFDFTLEGKEVAGE
YVGLPWPCKDHPGSPVLYNTDIEVAKGGMFRNFEGLVEGESLLAKNAPLNSPID
TGYPOITKDNEKVLGITLSAQEKEKMGSTWSDDSNIIATKCIKGIIVPYGNAKARA
VWTFKDKIPLHREPLHSRNDLVQKYPSEDKQALRYVDTKFSVQQAQDKYKEPFL
NLVTRVLNVLNGAGMENRSMILTRLPEMFEINPELAREQDIIKAGDMIIVHSPEPT
KIHVRVNPGVAKDMIFLPFHTGVNMQVDLTHNFEPTKPYASGESANTVTNYGYD
IMQIPETKGLCRISKDKG"
/complement(225222. .225359)
/gene="fdhA"
/note="Pfam match to entry PF00384 molybdopterin,
Prokaryotic molybdopterin oxidoreductases, score 32.60,
E-value 3e-09"
complement(226338. .226340)
/gene="fdhA"
/note="TGA codes for Selenocysteine in Cj1511c"
complement(226665. .226718)
/gene="fdhA"
/note="PS00551 Prokaryotic molybdopterin oxidoreductases
signature 1"
complement(226880. .227071)
/gene="Cj1513c"
complement(226880. .227071)
/gene="Cj1513c"
/note="Cj1513c, possible periplasmic protein, len: 63 aa;
no Hp match. Contains possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="possible periplasmic protein"
/protein_id="CAB73933.1"
/db_xref="GI:6968937"
/translation="MKNRRFLKKSAPALGAAGVLGGTTTLAKADERKDLVKGSKK
KEVLFORSANWENYIKAE"
complement(227046. .227759)
/gene="Cj1514c"
complement(227046. .227759)
/gene="Cj1514c"
/note="Cj1514c, unknown, len: 237 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1514c"
/protein_id="CAB73934.1"
/db_xref="GI:6968938"
/translation="MGALMDLEKILRLARSILYOCGLGELFVFSFEERLSKLOEVLKTM
OECLFDENLKSNEIDLKHLDDENSIOAFFKEYVDLLFLSKNSIPTFFSVIEEGFENS
NPLLCVRQIILSKIRNRKFRFESDSVGCFLLMSEFLRQNEDDUAKSELFKRVINK
SIDIEGLDFVPMKNANLYKEIASIAFMEFERLCFEVEKPAKINSKKVQNDLSRSEF
LRREANKRRTREKSGQIS"
complement(228063. .229211)
/gene="Cj1515c"
complement(228063. .229211)
/gene="Cj1515c"
/note="Cj1515c, probable decarboxylase, len: 382 aa;
similar to TR:Q56575 (EMBL:D31783) Vibrio alginolyticus
carboxynorspermidine decarboxylase (377 aa), fasta scores;
opt: 667 z-score: 807.5 E(): 0, 36.6% identity in 380 aa

overlap. 45.8% identity to HP0020"
/codon_start=1
/transl_table=11
/product="putative decarboxylase"
/protein_id="CAB73935.1"
/db_xref="GI:6968939"
/translation="MFYEKIQTPAYIIEEDLKRKNCELLASVGEKSGAKVILLAKGFA
FSGAMKIVGVLKGCCTGSLWEAKFAKYMDEKELHTYSPAFKDEDEIGEIASLSHHIVF
NSLAQHFHFOAKQKNSLGLRCNVEFSIAPKELYNPGRYSLRGIRAKDFENVDLSTI
EGLHFALCBESAGALEAVLKVFEKFGKWIQGMKWNFGGHHITKKGIDVEKLIAL
CKNFSDRKYGQVYLEPGEAGWQTNLVASVVDIIEENEKQIAIILDTSEAHMPTIIM
PBYHNARILATIRENEKISDLKESFSLTTLCTCLAGDMGEYAFDKKLKIGDKI
VFIDQIHTVTKNTFENGIRLPNMLLNAKNELOMIREFSYKDYSLRN"
/gene="Cj1516"
/complement(229335. .230876)
/note="Cj1516", probable periplasmic oxidoreductase, len:
1539 aa; similar to members of the multicopper oxidase
family e.g. YACK-ECOLI probable blue-copper protein yack
precursor (516 aa), fasta scores: opt: 376 z-score: 429.0
E(): 1.4e-16, 28.7% identity in 460 aa overlap, and
BLR0_MTRVE bilirubin oxidase precursor (572 aa), fasta
scores; opt: 309 z-score: 352.2 E(): 2.7e-12, 24.8%
identity in 499 aa overlap. No Hp match. Contains possible
N-terminal signal sequence, fam match to entry PF00394
Cu-oxidase, Multicopper oxidase, and PS00080 Multicopper
oxidases signature 2"
/codon_start=1
/transl_table=11
/product="putative periplasmic oxidoreductase"
/protein_id="CAB73936.1"
/db_xref="GI:6968940"
/translation="MNRNRLKFNALFLASMGVAYANPMIDMHSMHKNHNSHDLDTIS
FINFAPKNLLDPKQPPQGEILKALPLIKNESKEKNIFRATLEIKENHIELIKGKT
LEETVINGLVPAKTEVFGDKLEILVKNLKEATTIHGWVVPDQDQSPHDPDLIAG
EERYIREIPQDSAGTYWYHPHYTASQVFMGLAGAFVIKAKKDALSHLKEKDLMI
SRLRLDENAOITPNNNLNDWLNREGFVLNGQPKIKLATNRIYRNATARYLIN
LDLOGAKFLLVGTGGILIEKAIYKEELFSPASVEVLIIDAPKDNKFLSAVYDRDK
MNVKESTNLFANLNKLEKLELPKNLKFPLEEPKEKEIIMSEDHMQHSMGK
SENELKALASMFELNGSYDLKRIIDSLKGVVVEDWIVINKSHMDHPFHITHQFOFEL
ISSKLNKGKQKAEFRAPRDTINRPNEELRLKMKQDFKGLRMVHCHILEHEDLGMGN
LEVKE"
/misc_feature
229533. .229946
/gene="Cj1516"
/note="Pfam match to entry PF00394 Cu-oxidase, Multicopper
oxidase, score 26.50, E-value 0.0002"
230814. .230849
/gene="Cj1516"
/note="PS00080 Multicopper oxidases signature 2"
230878. .231099
/gene="moaD"
230878. .231099
/gene="moaD"
/note="Cj1517, moaD, possible molybdopterin converting
factor, subunit 1, len: 73 aa; similar to many predicted
moaD genes, e.g. TR:O31706 (EMBL:AF012285) Bacillus
subtilis molybdopterin converting factor (subunit 1) (77
aa), fasta scores; opt: 111 z-score: 167.7 E(): 0.051,
30.5% identity in 59 aa overlap, and weakly similar to
MOAD-ECOLI molybdopterin [mpt] converting factor, subunit
1 (81 aa), fasta scores; opt: 81 z-score: 127.5 E(): 8.8
38.5% identity in 26 aa overlap. 61.8% identity to HP0801"
/codon_start=1
/transl_table=11
/product="possible molybdopterin converting factor,
subunit 1"
/protein_id="CAB73937.1"
/db_xref="GI:6968941"
/translation="MVKVEFLGPINKENLELVKNLKEILQKDSLEKWELECAV
SLANDEIFNENTKIKGDKIALLPPVCGG"
231100. .231546
/gene="moaE"

/note="Cj1506c, probable MCP-type signal transduction protein, len: 700 aa; similar to many e.g. TR:CA83918 (EMBL:A132321) Halobacterium salinarum arginine chemotaxis CAR protein (452 aa), fasta scores; opt: 382 z-score: 382.7 E(): 5.4e-14, 29.9% identity in 281 aa overlap. Contains two possible transmembrane domains and pfam match to entry PF00015 MCPsignal, Methyl-accepting chemotaxis protein (MCP) signaling domain. 35.3% identity to HP0099"

/codon_start=1
/transl_table=11
/product="putative MCP-type signal transduction protein"
/protein_id="CA83927.1"
/db_xref="GI:6968931"
/translation="MFKSLNGLKLIPIFSAVAVVIGLVILSLITKQVSQNTKNTED ILASTRYATQTQIGFEGMIALNKSTGTTLMFRSTSKEDLDITNTITNTFDN SAYSNFTLYLIDPEPFKEESKFNTSGTKFVMLYADEEDKNGKIKAIQASDEIAN LQVQDILKRAKYGENKVIIGRPILKNLGGDFDANVAIPIFDKKNQVGVIGMTLD FSDIATYLLDPKQYDGLRLVNSDGFMAHPNKNLVLNKLDINPNKGAQETKYA ISEKNGVFNVIASGDSDSYAANSFKVQDSSWAVLVTPAKYSFKPLKLQILILGA SFPIFVVLGVVYCVKRVASRLPVLSSLESFEFLNHKEIKPKAIEIRANDELGA MGRINIENIEKIQISLEODQNAVDESQVTAIEIEKGNLTARITKNIPQILKMKLLA NRMLDVLQSKTGSNNNEINRVDSYKALDFSTEFVNAKGEVEITNIGKRIKKMLLA SSNFAKDLANGSEELKNSMQKADGSNAQSSLEQSAQAAVEEINSSMONVSGKTVEVA SOADDIKNIVNVIKDIAEOTNLALNAIAEARAAGEHGRGFAVVADEVROLAERTGKS LSEIEANILVQSVNEAVESVKQTAGITQINDAIAQLAETVTKRENVEVANVTNITN EVNGIAAILEDVNNKRP"

complement(218825..219007)
/gene="Cj1506c"
/note="Pfam match to entry PF00015 MCPsignal, Methyl-accepting chemotaxis protein (MCP) signaling domain"

complement(220830..221567)
/gene="Cj1507c"
complement(220830..221567)
/gene="Cj1507c"
/note="Cj1507c, unknown, len: 245 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1507c"
/protein_id="CA83928.1"
/db_xref="GI:6968932"
/translation="MDKNQKEIISYMRLEINSNEKLDGCTPARKIANKFNINIEKIG LADENHMRIDNCELGQFHLDFEAKITVLKKIIPSDIEXPRIFCDKARDIAKEGCG LKSMRSALKTKVDVYKQYQLGCFREKKGQFIVRTKWIENADGDLFGKRGTELELI QGTGSLHASKLMGINTKAKWMLHQTQKNSQELVSTROGRSKESGTKLTPRAMELM ENYSILOKDIIEYANKRFELFLKGRK"

complement(221560..222342)
/gene="fdhB"
complement(221560..222342)
/gene="fdhD"
/note="Cj1508c, fdhD, FdhD protein (required for formate dehydrogenase activity), len: 260 aa; similar to e.g. FDHD_WOLSU FDHD protein (286 aa), fasta scores; opt: 622 z-score: 728.9 E(): 0.385% identity in 252 aa overlap, and FDHD_ECOLI FDHD protein (277 aa), fasta scores; opt: 306 z-score: 363.8 E(): 6.1e-13, 25.0% identity in 244 aa overlap. No Hp match"

/codon_start=1
/transl_table=11
/product="FdhD protein"
/protein_id="CA83929.1"
/db_xref="GI:6968933"
/translation="MDPLFTTQILKYGDSEFTCNDFLVREIKLEIFINDEKVGALMA TPVDEQALAIGLMSENIEIAKVDIESIEFTKDDGMSVHIKAKIDKENLAKLNAGVVI SGCGRAHTANIDPEAIEASKISSDAKFSKDEILKQMSFTVTCLELYEKTGCVHTAKLF VDEKTFITGEDIAQHNITDKALGKACLAGLDSCFLMVSRLSENVAKAVMHKIPV LVSRPTCLGVMTARKFDLTLCGFARENKINITYSGEFLHG"

complement(222509..223441)
/gene="fdhC"
complement(222509..223441)
/gene="fdhC"
/EC_number="1.2.1.2"

/note="Cj1509c, fdhC, probable formate dehydrogenase, cytochrome B subunit, len: 310 aa; similar to e.g. TR:P28180 (EMBL:X54057) Wolinella succinogenes formate dehydrogenase C precursor (306 aa), fasta scores; opt: 634 z-score: 759.3 E(): 0.428% identity in 318 aa overlap, and FDNL_ECOLI formate dehydrogenase, nitrate-inducible, cytochrome B556(fdn) subunit (217 aa), fasta scores; opt: 189 z-score: 234.3 E(): 9.9e-06, 25.5% identity in 212 aa overlap. No Hp ortholog"

/codon_start=1
/transl_table=11
/product="putative formate dehydrogenase, cytochrome B subunit"
/protein_id="CA873930.1"
/db_xref="GI:6968934"
/translation="MRKVFTVLLSVLSFAYGSRMGQDTQIMDFHRTINIPNYDTF GKLMTLOGEVIAITIALIATVLSAFALHYVMIGPKQFSDHGKXIYAFILPERLRFH IAATSWILVLTGFMVMEGEVGGFVRYCKNLHAFATILFISIIIPMFLCWIKRML PASTDIRMMIVGGYLSKIKRPVAGKFNFGKSWIYIAVFGGFLMITITGGFMFLDF NSTAIQGLFGTQITELLRSIAIVHNLGICAVFGVHYIYMAVFAIKGSIHMSISYK EEEVYIILHSYWKLSNKKQIEPSFSYDPNVKI"

complement(223438..224079)
/gene="fdhB"
complement(223438..224079)
/gene="fdhB"
/EC_number="1.2.1.2"
/note="Cj1510c, fdhB, probable formate dehydrogenase iron-sulfur subunit, len: 213 aa; similar to e.g. FDHB_WOLSU formate dehydrogenase iron-sulfur subunit (EC 1.2.1.2) (200 aa), fasta scores; opt: 1025 z-score: 1153.1 E(): 0.695% identity in 197 aa overlap. No Hp ortholog. Contains PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature, and Pfam match to entry PF00037 fer4, 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains"

/codon_start=1
/transl_table=11
/product="putative formate dehydrogenase iron-sulfur subunit"
/protein_id="CA873931.1"
/db_xref="GI:6968935"
/translation="MSKVNFANLEKERLKFCDNERCIDNCGCAVACDEAHLEPHIR RRRVITLNEGQKEVSTISCMCHDDAPCSIVGPDGFIYRAGDGIYHDEIKGICGG YCLTACFPAGQFPKDSVFGNKGIMDKRCATNGGPEATNSERELIYQGNRIAGKVP VCAAMCTKALLVGESSKIEIYHNRLNRRNYGIPNPSESLKWIATYTKERL"

complement(223732..223905)
/gene="fdhB"
/note="Pfam match to entry PF00037 fer4, 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.. score 62.50, E-value 7.3e-17"
complement(223753..223788)
/gene="fdhB"
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding region signature"
complement(224076..226880)
/gene="fdhA"
complement(224076..226880)
/gene="fdhA"
/EC_number="1.2.1.2"
/note="Cj1511c, fdhA, probable formate dehydrogenase large subunit, len: 934 aa; similar to e.g. TR:P28179 (EMBL:X54057) Wolinella succinogenes formate dehydrogenase large subunit precursor (EC 1.2.1.2) (939 aa), fasta scores; opt: 3368 z-score: 3913.9 E(): 7.3e-211, 52.8% identity in 930 aa overlap, and FDHF_ECOLI formate dehydrogenase H (EC 1.2.1.2) (formate-hydrogen-lyase-linked, selenocysteine-containing polypeptide) (715 aa), fasta scores; opt: 813 z-score: 943.1 E(): 2.2e-45, 30.9% identity in 696 aa overlap. No Hp match. Contains pfam match to entry PF00384 molybdopterin, Prokaryotic molybdopterin oxidoreductases, and PS00551 Prokaryotic molybdopterin oxidoreductases signature 1. Note that TGA codes for Selenocysteine at

gene
complement(211197. .213284)
/gene="putp"
CDS
complement(211197. .213284)
/gene="putP"
/note="Cj1502c, putP, probable sodium/proline symporter, len: 495 aa; similar to many e.g. PUTP_ECOLI sodium/proline symporter (proline permease) (502 aa), fasta scores; opt: 1695 z-score: 1986.4 E(): 0, 51.7% identity in 487 aa overlap, 74.9% identity to HP0055. Contains Pfam match to entry PF00474 SSF, Sodium:solute symporter family"
/codon_start=1
/transl_table=11
/product="sodium/proline symporter"
/protein_id="CAB73923.1"
/db_xref="GI:6968927"
/translation="MIKERLEMEVVQINTPIAVMFVVALLMLYIGFTFYKQKNSESDY FLGRMGPIVISALSAGSDMSGLMLGPGALYVSGFIDSYIAIGTITGASLNWIFV AKRLIYTVIANSLTIPDYFETFRDDKHILRIVCAVILIFFTFVSSGLVSGAKL FSTFGIRYDALTGTITIIIVAYTFLGKYKAVCTWMIQGLMMMALIIVPLVMLYHL GGFGEAMNIVREIKPOALSMGEGVIVSIISALAWGLGYGOPHILVRPMSIBSTKDI PMATFGIAMAVCLLSACMIGILGIAYVHKFELSQDPEKFIIVMSQLLFNPWIAI LLSAILAAMSTASSQLLVSSSTIAEDFYKKIFREDAPSHIVNLGKLGVLVVAVIAF LIGTDKNSVLSIVYAWAGFGASFGSVMLFSFWRMTRVGAILGMITGAATVVLWK NFANSGLIETVPGFLVASVVIILIASLFTNVRSCTKAAYEKMLKEF"
complement(211959. .213155)
/gene="putp"
/note="Pfam match to entry PF00474 SSF, Sodium:solute symporter family, score 409.80, E-value 2.6e-119"
complement(213284. .216772)
/gene="putA"
CDS
complement(213284. .216772)
/gene="putA"
/EC_number="1.5.99.8"
/EC_number="1.5.1.12"
/note="Cj1503c, putA, possible proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase, len: 1162 aa; similar to e.g. PUTA_ECOLI bifunctional PUTA protein [includes: proline dehydrogenase (EC 1.5.99.8); delta-1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12) (1320 aa), fasta scores; opt: 611 z-score: 639.8 E(): 2.6e-28, 26.9% identity in 892 aa overlap, 65.0% identity to HP0056. Contains Pfam match to entry PF00171 aldedh, Aldehyde dehydrogenase, PS00017 ATP/Gnp-binding site motif A (P-loop), PS00687 Aldehyde dehydrogenases glutamic acid active site, and PS00070 Aldehyde dehydrogenases cysteine active site"
/codon_start=1
/transl_table=11
/product="putative proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase"
/protein_id="CAB73924.1"
/db_xref="GI:6968928"
/translation="MIQKALAAEELQGOATEANSISEKEFHAKMQKLLNNPKNVML IELDRSFRCKDNKNSFELIETLSKYGTADFFSTFEKFLFSLNFGFAPNLSPVF FVAKHREDTKAMVLDAMPVLEPHIKRKEDDKITLNVNLIGEVLGEAEKSKYRMOKY EEALKSSTYITYSIKTITJFSQINIDFNYSKEEVKRLDYLALALEBEKKOGVSKF INDMEEFRDLETTVAFMESVAKFDIKAGIVLQATIPDSIKYLKKLFAFSKERVUKG MKSIKIRFVKGANNESEETIASQRWELPTFYKKIDTDSNYKMLDFFILEGDNKYIN IGTAHNEIFEIAYATRISEVGALDSFTFEMLEGMSLQCSYELSKMHDLILYTPQGE IGHNNIAIYVRLRDENTSDNPMFYFNFLKVGDKNNIQQELFLKSLGKITLDNWT HRQDKDNESNIISSYEGKKFSNESPTDFILAQRVYAKEIKTKYENLKDYDVPVIG ELDFKAENLVLEVKKIEDRVIGKAYLAGAKEIKQALEVAKNSKSTQKSHDIYQIL AKSAKIRARPGRDLIGLALEVGKTEFLEIDPEVSEAIDEFIEPHSEELKKNPKVT FPKGIGVTIPWNPFGISVGTITAAPLAAGNVIYKPSLSLTGTMLCKCFWDSGI PKDALLIFPSKSDSIKYLILDEAIKFSILTGEDTYAYAMKANPITLLISAETGKRNA TIVSKFADRDASAIKNIHSAFNSGOKCSATSLVLEEYVHDEEFKTLVDAASMA VGNPEEFKNGLTADKPPSKVKOKALDELQPYEEMALKPKFLEDNPLYMTPTGKYCTK KGDTFHNELFVPILSYMRKADLKEAIDIVNSTGYGLTAGFESLDEREWEYFHTHIEA GNYIINKPTTGAIYLRQPFQGGIKKSAIGKRVGIYNIITQFLEIKROEDFDDHLIEDK FSKKQAIDLKDAQDFINLKNQSAYHAYHNEFSKDYVNIREDGNLFSYTRIKN MVLRLCCNESLKDVLGVLGANTIANIDLSISYDEQDVSMIERIVQSIGAKVILFKENK

ENFIKSIKEYRVRYLAKPDVNDIYKEAAKLAKIIIREKPLLNGRFELLNYFNEKAL SISFHRVGNLGIKRAIS"
complement(213881. .215242)
/gene="putA"
/note="Pfam match to entry PF00171 aldedh, Aldehyde dehydrogenase, score 127.30, E-value 2.8e-34"
complement(214457. .214492)
/gene="putA"
/note="PS00070 Aldehyde dehydrogenases cysteine active site"
complement(214553. .214576)
/gene="putA"
/note="PS00687 Aldehyde dehydrogenases glutamic acid active site"
complement(215000. .215023)
/gene="putA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(216903. .217829)
/gene="seld"
complement(216903. .217829)
/gene="seld"
/EC_number="2.7.9.3"
/note="Cj1504c, seld, probable selenide,water dikinase, len: 308 aa; similar to e.g. SELD_ECOLI selenide,water dikinase (selenophosphate synthetase) (EC 2.7.9.3) (347 aa), fasta scores; opt: 638 z-score: 720.3 E(): 8.4e-33, 37.2% identity in 298 aa overlap, No Hp match. Contains Pfam match to entry PF00586 AIRS, AIR synthase related proteins"
/codon_start=1
/transl_table=11
/product="putative selenide,water dikinase"
/protein_id="CAB73925.1"
/db_xref="GI:6968929"
/translation="MOKTALLSDIGNEDASVYQISPDIALVQTDLTFTIPVIDSAYH FGAIAAANLSDVFMAGAEVINALNIIVGFTDCHNLDNLKELLEKANDKVKCNALVV GGHITESTLFFGLSVTGVKHPSTANTSKIGDCIILTKPLGTGTLSTALKAQMLN QKHLIDMLKNMIELNYSQAIALKFHPSAMSDVTGFLGLHLKEMLNKNTSFEIFESE LPLDGVKVEYFNNGLIPAGAYKNKEIKELTDPDNEEKLILCDPQPSGGLLIISEKD SLECLKEDENIQAKIIIAIKVKNQENDIIIS"
complement(216999. .217826)
/gene="seld"
/note="Pfam match to entry PF00586 AIRS, AIR synthase related proteins"
complement(217935. .218507)
/gene="Cj1505c"
complement(217935. .218507)
/note="Cj1505c"
/note="Cj1505c, unknown, len: 190 aa; similar in N-term to hypothetical proteins e.g. Y990_METUA (75 aa), fasta scores; opt: 121 z-score: 164.1 E(): 0.081, 33.3% identity in 69 aa overlap, No Hp match. Some similarity to Cj1492c (21.1% identity in 190 aa overlap). Contains pfam match to entry PF01206 UPF0033, Uncharacterized protein family UPF0033"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1505c"
/protein_id="CAB73926.1"
/db_xref="GI:6968930"
/translation="MKIDCRNLSCPOPIVETKNALEKLOENEILEVLNIISSKNNVV KFLNSLNPVNDENAOEFCIKVQKNFSSSEVNIHDYVNLTKTKVGGELGQNLIL GFSLSTLKNLDHAPSILCVNESVLNVNDENHKAHLENLGIEIISCGACLEFF NKSKEIKIGNAYEILNELFGKAKIIL"
complement(218301. .218507)
/gene="Cj1505c"
/note="Pfam match to entry PF01206 UPF0033, Uncharacterized protein family UPF0033, score 18.10, E-value 0.00023"
complement(218618. .220720)
/gene="Cj1506c"
complement(218618. .220720)
/gene="Cj1506c"
/gene="Cj1506c"

| | | | |
|--------------|---|--------------|--|
| misc_feature | complement(206765..207241) /gene="carA" | misc_feature | /translation="MSKADIIVGIQWDEGKGVVVKLCENYDFVCRSAGGHNAGHTI WNGVRYALHMLPSCVGLHPCINIIINGVVSPEVLIETAMQFENGLRGLYISDRALH NLKHSILDIKAEKLGKNAITGTCGIGPSYADKINRTHRGVGLLEDEKRLCEALIK DFEANKTFEEMLEIEIPSAEELLADLRFNELLPTITDTTMLMKALDEDKRVLLG AQGMLDIDHGTYPYVSSSTISAGTLTGIGLNPKAGNIIGIVKAVATRVNGAFPT DKGEDGEKIAQIGKEIGVSTGRKRCGWPDAVAVRYTARLNGDALSMLKLDVLDGF EKIKICRAYEYKGMEDYIPSDLENVQPIYEMDGNKVFQIKDYKIDLLPENAKYIAR LEELAGVKYIISTSPERDDTIIL" |
| | complement(207247..207798) /gene="Cj1495c" /note="Cj1495c, unknown, len: 183 aa; similar to hypothetical proteins e.g. TR:067633 (EMBL:AE000754) Aquifex aeolicus AQ_1752 (183 aa), fasta scores; opt: 268 z-score: 310.8 E(): 5.4e-10, 24.6% identity in 183 aa overlap, 45.4% identity to HP1236" | | complement(208923..210062) /gene="purA" /note="Pfam match to entry PF00709 Adenylsucc_synt, Adenylsuccinate synthetase, score 613.50, E-value 1.2e-180" |
| CDS | complement(207247..207798) /gene="Cj1495c" | misc_feature | complement(209658..209693) /gene="purA" /note="PS00513 Adenylsuccinate synthetase active site" |
| | complement(207891..208409) /gene="Cj1496c" | | complement(210024..210047) /gene="purA" /note="PS01266 Adenylsuccinate synthetase GTP-binding site" |
| gene | complement(207891..208409) /gene="Cj1496c" | trna | complement(210105..210202) /gene="trNA-Sec" |
| | complement(207891..208409) /gene="Cj1496c" | | /note="trNA Sec anticodon TCA, Cove score 35.94" |
| CDS | complement(207891..208409) /gene="Cj1496c" | gene | complement(210105..210202) /gene="trNA-Sec" |
| | complement(207891..208409) /gene="Cj1496c" | | 210342..211550 /gene="Cj1500" |
| misc_feature | complement(207891..208409) /gene="Cj1496c" | CDS | 210342..211550 /gene="Cj1500" |
| | complement(207891..208409) /gene="Cj1496c" | | 402 aa; similar to hypothetical integral membrane proteins e.g. YEDE_ECOLI (401 aa), fasta scores; opt: 722 z-score: 819.1 E(): 0, 44.3% identity in 395 aa overlap. No Hp match" |
| misc_feature | complement(207891..208409) /gene="Cj1496c" | misc_feature | /codon_start=1 /transl_table=1 /product="putative integral membrane protein" |
| | complement(207891..208409) /gene="Cj1496c" | | /protein_id="CAB73921.1" /db_xref="GI:6968925" |
| CDS | complement(207891..208409) /gene="Cj1496c" | gene | /translation="MNSFKQKYLINFDNRSKSMIALGILSAVYFGIFGVWAVTGEMT RWGSEFLELLGMNDIGYSYQKQNLGTLRTDGIIMLIGMFIQGLVAALIANKVKFR LPASNIRIFOAIVGIGILSGYARLAFGCNLANPFTGLPYFSLHTWLTVMVLGIYLG VKICNTSFEPKAKLERVKNENPLNKOSLRTKLYFNGLILLFTAFVWVFLVFTNG NISTONKQSLALALHIFGVFGFVISRQICFTSCFRDLFLFGRDNRKALIGMIIA SLTAFALIQHTSKLIELSPAVAVGAPLFGVIFAGGCGCGWTYAFEGQSHFIIV GIANIIGTILALSYNLPLKAFKEGIKINLITFEGLNGFFINLILFILMFVFFVYFK KHFFKNLKG" |
| | complement(207891..208409) /gene="Cj1496c" | | 211554..211781 /gene="Cj1501" |
| CDS | complement(207891..208409) /gene="Cj1496c" | CDS | 211554..211781 /gene="Cj1501" |
| | complement(207891..208409) /gene="Cj1496c" | | /note="Cj1501, unknown, len: 85 aa; similar to hypothetical proteins e.g. YEDF_ECOLI (77 aa), fasta scores; opt: 341 z-score: 453.8 E(): 5.9e-18, 65.7% identity in 70 aa overlap. No Hp match. Contains PS01148 Uncharacterized protein family UPF0033 signature, and Pfam match to entry PF01206 UPF0033, Uncharacterized protein family UPF0033" |
| misc_feature | complement(207891..208409) /gene="Cj1496c" | misc_feature | /codon_start=1 /transl_table=1 /product="hypothetical protein Cj1501" |
| | complement(207891..208409) /gene="Cj1496c" | | /protein_id="CAB73922.1" /db_xref="GI:6968926" |
| CDS | complement(207891..208409) /gene="Cj1496c" | misc_feature | /translation="MKITYSLNLQGEACYPYPAIATLDVLPKLSGEILEVLCDCPOSI NSTPQAKNRGVLEIDQSGPTLRLFIQPK" |
| | complement(207891..208409) /gene="Cj1496c" | | 211560..211775 /gene="Cj1501" |
| CDS | complement(207891..208409) /gene="Cj1496c" | misc_feature | /note="Pfam match to entry PF01206 UPF0033, Uncharacterized protein family UPF0033, score 101.00, E-value 2.3e-26" |
| | complement(207891..208409) /gene="Cj1496c" | | 211572..211610 /gene="Cj1501" |
| CDS | complement(207891..208409) /gene="Cj1496c" | misc_feature | /note="PS01148 Uncharacterized protein family UPF0033 signature" |
| | complement(207891..208409) /gene="Cj1496c" | | |

misc_feature
PDLARVGNVYRTADWHEHNMWDPTSVVPDSIMPSYKHKMFKNNADETAYAEALTVKKVF
NVPIYDRENGTKLGSWEDAQAEVRAEQAIIVDMKNQEVKDAFAKEIKEIVALIAYLN
SLK"
complement(201780. .201797)
/gene="ccoO"
/note="PS00190 Cytochrome c family heme-binding site
signature"
complement(201990. .203456)
/gene="ccoN"
complement(201990. .203456)
/gene="ccoN"
/EC_number="1.9.3.1"
/note="Cj1490c, ccoN, cb-type cytochrome C oxidase subunit
I, len: 488 aa; similar to TR:087193 (EMBL:AB018105) H.
pylori characterised cb-type cytochrome C oxidase subunit
I (EC 1.9.3.1) (488 aa), fasta scores; opt: 2471 z-score:
2791.9 E(): 0, 68.0% identity in 487 aa overlap (-HP0144,
68.8% identity). Contains PS00077 Heme-copper oxidase
catalytic subunit, copper B binding region signature, and
pfam match to entry PF00115 COX1, Cytochrome C and Quinol
oxidase polypeptide I"
/codon_start=1
/transl_table=11
/product="cb-type cytochrome C oxidase subunit I"
/protein_id="CAB73912.1"
/db_xref="GI:6968916"
/translation="MHFGNVLNYDVTYARYFMFATILFGIVGMAICTLIAFQWAPNL
NYLAGEYAFPSRLPLHTSCVIFGFMLSGIWATYVIGQVLKVSMAESRFLMVGKL
HFVLMITWLVAVISFPMGVTTSKEYAELEWDLIDLVLMVGVSIFGLGIRREK
TLVISLWYIATELGIAMLYLFNNMEVPTYFTVMGKWHWSVSMYAGTNDALVQWRYG
HNAAVFVFGIIAQIYFLPKESQGIIFS KLSLFAFWGLMFVYLWAGGHLLIYSTV
PDMQWQTSVFSVLLIPSGSAINILLFMKGWSQLSRESPLIKFWILASTFYMFSTL
EGPILSKVNALAHFTDMIPGHVHDSGLGWGFMFMAALYHMTPRVFKRELYSKSLM
EAQFWIQTGIVLYFASMIAGITQGMWRATDEYGNLLYSFIDTVVAIVPYWIRAI
GGULIGIFMFTYNIYKTIACGRVLDKEPKSASPMAA"
complement(2021128. .203096)
/gene="ccoN"
/note="Pfam match to entry PF00115 COX1, Cytochrome C and
Quinol oxidase polypeptide I, score 107.70, E-value
3.6e-29"
complement(202647. .202814)
/gene="ccoN"
/note="PS00077 Heme-copper oxidase catalytic subunit,
copper B binding region signature"
complement(203582. .204262)
/gene="Cj1491c"
complement(203582. .204262)
/gene="Cj1491c"
/note="Cj1491c, probable two-component regulator, len: 226
aa; similar to many e.g. VANR_ENTFA regulatory protein
VANRB (220 aa), fasta scores; opt: 322 z-score: 369.3 E():
3e-13, 29.9% identity in 214 aa overlap. No Hp ortholog.
Contains Pfam match to entry PF00072 response_reg,
Response regulator receiver domain"
/codon_start=1
/transl_table=11
/product="putative two-component regulator"
/protein_id="CAB73913.1"
/db_xref="GI:6968917"
/translation="MSQECKELIILVEDEVKARESMINILSRFSKVLGAQNGDGL
KFKFKPLVDITDIAMPDGLDMAREIKEISDDVPIVVLVSAYSKERLLSIDIGI
DKYLIKPDVIEELFVLDLIGEKIEANMLVKISEQYQFNKRTLIYSGGEIVLTKK
ELAFISLLKPGALVLEDIKKNVWIGHVSVDATVRTFIKVRDKVGDFFIKNPVSL
GKYNINK"
complement(203906. .204241)
/gene="Cj1491c"
/note="Pfam match to entry PF00072 response_reg, Response
regulator receiver domain, score 116.80, E-value 3.9e-31"
complement(204265. .205476)
/gene="Cj1492c"
complement(204265. .205476)
/gene="Cj1492c"
/note="Cj1492c, probable two-component sensor, len: 403

misc_feature
aa; similar to many e.g. KINC_BACSU sporulation kinase C
(428 aa), fasta scores; opt: 281 z-score: 301.1 E():
1.9e-09, 29.3% identity in 242 aa overlap. No Hp ortholog.
Contains Pfam match to entry PF00512 signal, Signal
carboxyl-terminal domain"
/codon_start=1
/transl_table=11
/product="putative two-component sensor"
/protein_id="CAB73914.1"
/db_xref="GI:6968918"
/translation="MKDFFKQDFKALEKNTIFSRADVQGNLIPVSKLQICISYSKK
ELTGKHSIFKHPDVEECYIEELLKLSYKPKYVIFKNIDKLGKTYLETLIPLILD
KNNELTEIVAFSHDVNSFKLNEELANLHAKLRELSINLENTVKNHQOEIQLGKKE
KMQOALEKNEKDIKAYEILKSSLEQMSIDIAHOWRQPLNELGIAMFOMKONLKOE
KGFAEIVSOSKDMIKNMSETIDVFTLFNKGVCQVFIKETLANKALEIAFETIEKNHV
NINIVSKSDYELAYENGLIRVFLNLLINSIEAFKNNKKRIITINFSKFGKNNLKITI
KNWAGIDINDLKIIFQPYFTTHPSQGVGVYIISRIIESFQGIKIVKNGKDGACF
EIFLKLKERVE"
complement(204286. .204894)
/gene="Cj1492c"
/note="Pfam match to entry PF00512 signal, Signal
carboxyl-terminal domain, score 76.10, E-value 7.5e-19"
complement(205473. .206132)
/gene="Cj1493c"
complement(205473. .206132)
/gene="Cj1493c"
/note="Cj1493c, probable integral membrane protein, len:
219 aa; 33.9% identity to HP0861"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73915.1"
/db_xref="GI:6968919"
/translation="MSIDFLAIIISVAFLSFSGHCYSMCGGNLLFLINLSKSNLFL
TEIYNLFRIFAYIIVGIIIFGTGNILAINAKVOSLSFFILGIPFMILGFAIIFRGNML
SFENNVFDFSFKIIKKSNFKGLKSAIFLGNSGVPCLGVYFFFIANMSKQNI
EGILVMNIFGISTLPAMLFESKISQFSDFLKNLNYLSGVIVCYGINLAYIGFKAF
Q"
complement(206129. .207247)
/gene="carA"
complement(206129. .207247)
/gene="carA"
/EC_number="6.3.5.5"
/note="Cj1494c, carA, probable carbamoyl-phosphate
synthase small chain, len: 371 aa; similar to many e.g.
CARA_ECOLI carbamoyl-phosphate synthase small chain (EC
6.3.5.5) (382 aa), fasta scores; opt: 946 z-score: 1094.4
E(): 0, 40.3% identity in 377 aa overlap. 57.8% identity
to HP1237. Contains PS00442 Glutamine amidotransferases
class-I active site, Pfam match to entry PF00117 GATase,
Glutamine amidotransferases class-I, and Pfam match to
entry PF00988 CPSase_sm_chain, Carbamoyl-phosphate
synthase small chain, CPSase domain"
/codon_start=1
/transl_table=11
/product="carbamoyl-phosphate synthase small chain"
/protein_id="CAB73916.1"
/db_xref="GI:6968920"
/translation="MKAYIILENDIFLSAKAFKSGTFFGELVFNTSLGYQBIISDP
SVAGQFIVFMPETGIVGTNENDNSKEIFASGLMRELSFSNFRAKESQDYLEK
HGKIGIYELTRYLVKMRNNGLRAVISTEISNEDKLIALEKSAKIDEVNFRKVS
TKNYSKQGVNNAFQKFNDSKQVADYGVKTNILNELVGVGEVEVYPYV
KADELITLYKGEIQGVLSNGPGEPILKQVIAETIKKLAERAKIPMLGICLGHOLLN
AFGYETKMKFGQHGANGHPVINDTKVETITAGNNHNYNYPELAQVAHATHRNLFQDN
VEGVRYKDYPIISVQHPSGPHESKIFIFEKFMMLM"
complement(206135. .206686)
/gene="carA"
/note="Pfam match to entry PF00117 GATase, Glutamine
amidotransferases class-I,
complement(206426. .206461)
/gene="carA"
/note="PS00442 Glutamine amidotransferases class-I active
site"

FTPKRVSNEFLNNEVRKKAGLSYDRRENLOREYVESILMKNALEVSICFVENEESK
SRFLELDFREYETHIHQKAYLNALKLDYSGIKENLPIKAPILKHNPFPEPLSF
FNLEKQRTYYRYIILNLAEPRLSESKAKQNGFIHKMLEIYYKNYANNDFDIK
FANLDRQYKINISELEDFVKLFQMPAKNEKEHFGYGVVHTELENNILKLT
DSIKUGTIDRIDSKEGNLIIDYKSGVPNSYOLAFYQALYDNASVGFYDLNSMQ
ILHOKASDELRELRKDLVLMSEKEIEFENEQDEYCPYKLIYKKEK"
/gene="Cj1483c"
/complement(198758..199279)
/gene="Cj1483c"
/complement(198758..199279)
/note="Cj1483c, probable lipoprotein, len: 173 aa; 30.8%
identity (in N-term) to HP0150. Contains probable
N-terminal signal sequence and appropriately positioned
PS00013 Prokaryotic membrane lipoprotein lipid attachment
site"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAB73905.1"
/db_xref="GI:6968909"
/translation="MLETKKSPVPGILLSLIIACIIVTIFTASHHPVEDDFYD
SYQVNNYNOIQOQVNFDFKVFONDKITFIGRKRPISEYVDQNSYVANKISA
LQNIINSDLKVQVLLTRPFTANFDQKLEGQVKNGLISLALPKLDKGRWELKIFYANQ
ETVGFYSYELNAQ"
complement(199208..199240)
/gene="Cj1483c"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(199266..199868)
/gene="Cj1484c"
complement(199266..199868)
/gene="Cj1484c"
/note="Cj1484c, probable membrane protein, len: 200 aa; no
Hp match. Contains probable N-terminal signal sequence and
probable C-terminal transmembrane domain"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73906.1"
/db_xref="GI:6968910"
/translation="MKKILQGFELAIFFVLLPSRFALDAVLFNENILSQKVSNEI
NLIKELYQKSNIFIGVWGDKTETILLNKKELPOSYILLLSKNSHKVDIVGSKG
ALALIDKEAVLSPYGTGSLIPILATNKGDIYNAAILNGYADIVDRVAKSLGLQLEHS
IGNANROTINILRIIYGFICFALLYAQRRIKRNVRN"
complement(199865..199966)
/gene="Cj1485c"
complement(199865..199966)
/gene="Cj1485c"
/note="Cj1485c, probable periplasmic protein, len: 33 aa;
no Hp match. Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73907.1"
/db_xref="GI:6968911"
/translation="MSKALEYLVIGLVAAAIATAWSLVTVNHLHG"
complement(199963..200184)
/gene="Cj1486c"
complement(199963..200184)
/gene="Cj1486c"
/note="Cj1486c, probable periplasmic protein, len: 73 aa;
no Hp match. Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73908.1"
/db_xref="GI:6968912"
/translation="MMENNRCSVLSGVTCMLTIATVLLLVILAVLTWGLKAQOQVM
QKPYSLKDVQSKWFGSKQDHRISKEAQ"
complement(200181..201044)
/gene="ccop"
complement(200181..201044)
/gene="ccop"

/EC_number="1.9.3.1"
/note="Cj1487c, ccoP, cb-type cytochrome C oxidase subunit
III, len: 287 aa; similar to TR:087196 (EMBL:AB018105) H.
pylori characterised cb-type cytochrome C oxidase subunit
III (EC 1.9.3.1) (292 aa), fasta scores; opt: 952 z-score:
1043.8 E(): 0, 51.0% identity in 292 aa overlap (-HP0147,
51.2% identity). Contains 2x Pfam match to entry PF00034
cytochrome_c, and 2x PS00190 Cytochrome c family
heme-binding site signature"
/codon_start=1
/transl_table=11
/product="cb-type cytochrome C oxidase subunit III"
/protein_id="CAB73909.1"
/db_xref="GI:6968913"
/translation="MOWLNEDNVNLLSILGAILLITLVIVGRMFQMKREKKGSE
LSEHSWDGIGEYKNAVPTGWVFFLTIVWAIWYFLGWYPLNSFGSIQYNEEVATHN
TKFEKFNLSPEDKIAMQNFIVQCSACHGITGDGNGAQNINWGSSEGINVI
KHGSKGMNPPGSGMIGAADLGAVEREDIPAIKAYVAKDLISATKKTANENLVAKGKAY
TCAACHGEDGKGQGMFPDLTKYGSAAFFVDVLHSGKAGFTGTMPSFTLNDIQEAV
GEYVLSLRGE"
complement(200193..200426)
/gene="ccop"
/note="Pfam match to entry PF00034 cytochrome_c,
Cytochrome c, score 20.80, E-value 0.0039"
complement(200370..200387)
/gene="ccop"
/note="PS00190 Cytochrome c family heme-binding site
signature"
complement(200451..200693)
/gene="ccop"
/note="Pfam match to entry PF00034 cytochrome_c,
Cytochrome c, score 12.70, E-value 0.025"
complement(200643..200660)
/gene="ccop"
/note="PS00190 Cytochrome c family heme-binding site
signature"
complement(201044..201307)
/gene="ccop"
complement(201044..201307)
/EC_number="1.9.3.1"
/note="Cj1488c, ccoQ, cb-type cytochrome C oxidase subunit
IV, len: 87 aa; similar to TR:087195 (EMBL:AB018105) H.
pylori characterised cb-type cytochrome C oxidase subunit
IV (EC 1.9.3.1) (72 aa), fasta scores; opt: 143 z-score:
226.7 E(): 2.6e-05, 35.1% identity in 57 aa overlap
(-HP0146)"
/codon_start=1
/transl_table=11
/product="cb-type cytochrome C oxidase subunit IV"
/protein_id="CAB73910.1"
/db_xref="GI:6968914"
/translation="MEHLSIVDFVIKLLITFDLATVOKHEWEIFQGYGFALVMFLSI
VLXAVWVHLRYTEKKGERNYENYANLAKDDIDDSVLESKRSA"
complement(201312..201977)
/gene="ccoo"
complement(201312..201977)
/gene="ccoo"
/EC_number="1.9.3.1"
/note="Cj1489c, ccoO, cb-type cytochrome C oxidase subunit
II, len: 221 aa; similar to TR:087194 (EMBL:AB018105) H.
pylori characterised cb-type cytochrome C oxidase subunit
II (EC 1.9.3.1) (232 aa), fasta scores; opt: 1010 z-score:
1158.4 E(): 0, 65.9% identity in 220 aa overlap (-HP0145,
66.8% identity). Contains PS00190 Cytochrome c family
heme-binding site signature"
/codon_start=1
/transl_table=11
/product="cb-type cytochrome C oxidase subunit II"
/protein_id="CAB73911.1"
/db_xref="GI:6968915"
/translation="MFSWLEKNPFFFAVAVFVVIYAGIVEVLNPFNAENARPISGKRP
YTVIQLAGRAVYIKDSCNACHSQLIRPKSETDRYGMYSVSGEFAYDRPFLWMSKRTG

```
/gene="Cj1477c"
/note="Pfam match to entry PF00702 Hydrolase, haloacid
dehalogenase-like hydrolase, score 56.30, E-value 6.7e-13"
complement(191648..192607)
/gene="cadF"
complement(191648..192607)
/gene="cadF"
/note="Cj1478c, cadF, outer membrane fibronectin-binding
protein, len: 319 aa; 82.8% identical to TR:O06895
(EMBL:U07559) C. jejuni CADF precursor
(fibronectin-binding protein) (326 aa), and similar to
many oprF proteins e.g. PORP_PSEFL outer membrane porin F
precursor (root adhesin) (326 aa), fasta scores; opt: 401
z-score: 458.6 E(): 3.2e-18, 29.1% identity in 316 aa
overlap. No Hp match. Contains PS01068 OmpA-like domain,
and Pfam match to entry PF00691 OmpA, OmpA family"
/codon_start=1
/transl_table=11
/product="outer membrane fibronectin-binding protein"
/protein_id="CAB73900.1"
/db_xref="GI:6968904"
/translacion="MKKIFLCGLASVLCGADNNVKEIPTLANNVFEGLMDNRY
APGRILGHEDFDWLDQLEGLHYSDVKYTNKTTDITRTYLSAIGDIVGKFPF
YLAGGYEDFSNAAAYDNKSGFGHYGAGVKFRLSDSLALRLETRDQINFNHNMY
STLGISFGGKKEAVEADTRATPAQKCPVEPRECALIDENGCKTISLEHGFGF
DKTTINPTFOBKIEIAKIVLDENERTDITLGHDTNIGSRAYNOKLSERRAKSVANEL
EKYGVKSRITKVCYGODNPRESSNDTKEGRADNRRYDAKFILR"
complement(191678..191965)
/gene="cadF"
/note="Pfam match to entry PF00691 OmpA, OmpA family,
score 133.50, E-value 3.9e-36"
complement(191732..191866)
/gene="cadF"
/note="PS01068 OmpA-like domain"
complement(192732..193121)
/gene="rpsI"
complement(192732..193121)
/gene="rpsI"
/note="Cj1479c, rpsI, 30S ribosomal protein S9, len: 129
aa; similar to many e.g. RS9_ECOLI 30S ribosomal protein
S9 (129 aa), fasta scores; opt: 400 z-score: 508.5 E():
5.3e-21, 49.6% identity in 125 aa overlap. 77.5% identity
to HP0083. Contains PS00360 Ribosomal protein S9
signature, and Pfam match to entry PF00380 S9, Ribosomal
protein S9"
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S9"
/protein_id="CAB73901.1"
/db_xref="GI:6968905"
/translacion="MATYATGKRKTAIAKVVWPKSGKISVNGVDLNTWLGHEAIK
LKVVQPLLVTKQETSMDKATTLGGYSQAQEARLHGISRALAAWDADFRALLPKGL
LTRDSRTVERKKYGRRKARSPQFSKR"
complement(192735..193100)
/gene="rpsI"
/note="Pfam match to entry PF00380 S9, Ribosomal protein
S9, score 165.80, E-value 9.6e-51"
complement(192864..192920)
/gene="rpsI"
/note="PS00360 Ribosomal protein S9 signature"
complement(193124..193549)
/gene="rplM"
complement(193124..193549)
/gene="rplM"
/note="Cj1480c, rplM, 50S ribosomal protein L13, len: 141
aa; similar to many e.g. RL13_ECOLI 50S ribosomal protein
L13 (142 aa), fasta scores; opt: 490 z-score: 616.2 E():
5.3e-27, 53.1% identity in 130 aa overlap. 61.9% identity
to HP0084. Contains PS00783 Ribosomal protein L13
signature, and Pfam match to entry PF00572 L13, Ribosomal
protein L13"
/codon_start=1
/transl_table=11
```

```
/product="50S ribosomal protein L13"
/protein_id="CAB73902.1"
/db_xref="GI:6968906"
/translacion="MTKITKPEVKREWIVLDAEGKRGRLTTEVATILRGKNKPCFT
PNVDCGYVILIINASKAVFTGANKAEDLYHRHSGYFGSVKSEKFGDLEENPAKLYK
LAVRGMLPKTNIGRAMLKKLIYAGSEPHPTAQIAKEGK"
complement(193133..193519)
/gene="rplM"
/note="Pfam match to entry PF00572 L13, Ribosomal protein
L13, score 192.10, E-value 9e-54"
complement(193169..193237)
/gene="rplM"
/note="PS00783 Ribosomal protein L13 signature"
complement(193633..196398)
/gene="Cj1481c"
complement(193633..196398)
/gene="Cj1481c"
/note="Cj1481c, possible helicase, len: 921 aa; weak
similarity to several predicted helicases e.g. UVRD_MYCGE
putative DNA helicase II homolog (703 aa), fasta scores;
opt: 220 z-score: 237.1 E(): 7e-06, 24.5% identity in 715
aa overlap, and to e.g. ADDA_BACSU ATP-dependent nuclease
subunit A (contains helicase activity) (1232 aa), fasta
scores; opt: 204 z-score: 216.2 E(): 0.0001, 23.0%
identity in 868 aa overlap. 30.5% identity to HP1553. Also
similar to predicted helicases Cj1101 (23.9% identity in
532 aa overlap), and Cj0777 (24.1% identity in 673 aa
overlap). Contains PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=11
/product="putative helicase"
/protein_id="CAB73903.1"
/db_xref="GI:6968907"
/translacion="MSQFEFPLALEASAGSKTALSVRVALILKGARINEILALFE
TKKAANEMQRIILTFNLNLEKNKSECNELCKLLGDKDEELISLRDAKKEFEFLTE
KISTFAFFGKILRVFALNLCSSDFTMSERDLVDREIFLKLKDLKDLAYINLV
DEKENFELEKFEYENAYFNRPKILNPTKAYINKAYSELRSYCLGLTHVKNYNLC
DEKSEVLDLSVFMQSSPFTKPESTKYLDQLESTNLHESAKRMETINALTYALELNY
KIANMLLNHYSEAKNIFHKNTLNFQDVSKKVELITSEFKMDIYFRUDGFIHL
LKBQFQTSVIQYQILRLPIALIELVSGGVKKNTFFVYVGDKKQSIYRFRKGKELFD
LIDFQSTIKSDNLNTYRSKRLVDVFNTEPFKEIKDYKEQFALESKGGFVRIVES
KQFVQNAQAEIKETKLTLETFQINFLRSKNSYDDICLCKWNSDADMVLDFLREONI
PAFTQSNVLENKASVRLVLEYAKYCFGDEFYLVFLKELGFRPKITLDFSNKME
NVFLRELKLDNDIALIQIEYAKFNKFLLLFEPKALKIYSEQNMGISIMSVHK
SKGLEFDHVLDDSLKNSNNDIMLEIDINQWQLHKDKIKELTKPEIYTLTFKEN
ITRANEDDINKLYVAFTRAKESLIIRKNEESVNGYPSYFGKGFNLNHSQERGFLE
SQEILSVKKESIQTLQKFEKITLQEQSERLDSKELYFGNAFFHFMQNLKLPKGEN
FOILTORCKSKFRHFLDESDFEKLFRKRIELLKNIQFNLIGDGLKLLKEQALSNGEI
KQLDLLAKDEAFIIDYKTGLAMQDKKKEQVGYKTAISEILQDKVRAFIYVCLEN
EQIILEI"
complement(196342..196365)
/gene="Cj1481c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(196395..198761)
/gene="Cj1482c"
complement(196395..198761)
/gene="Cj1482c"
/note="Cj1482c, unknown, len: 788 aa; 29.5% identity to
HP1089"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1482c"
/protein_id="CAB73904.1"
/db_xref="GI:6968908"
/translacion="MKLIRIFSSSRQIREYXNQKQNALDSAIHIGFELDKVCLSNF
HKASSYSELILMOEACLKDKLEKLGISVEFFAPFAKNNELVFFFKELSEKKSID
LKNNDYATVYNEHLEILDEYKYNVALLEKNSFYDDLSPKNYTLNKLDFLDEYAIIV
DLQGLSKFEENLLSEISOIKEVLSFKTNELLYLLKLDLFLKIPDLKINTHYEINL
SCQEILKEEFKTKNKKMLKSPFELRALQCAFVMEISHSVKRGKLEKNTAVITPDDES
FCSEFLRFLDKDNMLNFASGISIKESLPYQKFDALYESASFVYKNOEDYFTQMI
FDYHNTLLHSLKLDIFDEKPKYFEKILLFLLENKEQLMKELEYFIK
DLLKNQSLTLKELIHLFMQISQLSDVGGKVTVMGLLESRLGCLFDGVLVDNFE
```

```
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73894.1"
/db_xref="GI:6968898"
/translation="WOERIKELERKYFLKLYLLKYLILILISVIAFCFVLMOKY
NKQNIYQATIEHKHKLQKILQAOILOENKNIISREKLYKELEEVKAVQENPHISKIE
IDSKLINTSLDKSPYQNSYEKALNLAKKYFDIKAYOKTIWALKANELDKQKQDSW
LIFAQAKRALGEKEAQSALDAYINYGLMELDGR"
complement(185022..185630)
/gene="Cj1473c"
CDS
complement(185022..185630)
/gene="Cj1473c"
/translation="Cj1473c, probable ATP/GTP-binding protein, len: 202
aa; no Hp match. Contains PS00017 ATP/GTP-binding site
motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative ATP/GTP-binding protein"
/protein_id="CAB73895.1"
/db_xref="GI:6968899"
/translation="ASKLIPREEIFHOINQILLESQKAFIFLWKGSGKSVLLQRLA
KYNVDFINENFKQDSFLKEKIEFLISQGSLLIILDEVGMVDYAMLESIRIYSDSISF
VLSSHNLINILKKEHFKSRLKACFKLNISLLEDDYIKLFGMNFNKNKCLKFLQKIS
QGNRYIDTKLSYEYINSFFDNKKSQEYILKLSALENGLLR"
complement(185520..185543)
/gene="Cj1473c"
misc_feature
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(185623..187041)
/gene="Cj1474c"
CDS
complement(185623..187041)
/gene="Cj1474c"
/translation="Cj1474c, probable type II protein secretion system
D protein, len: 472 aa; similar to e.g. GSPD_PSEAE general
secretion pathway protein D precursor (658 aa), fasta
scores; opt: 305 z-score: 332.5 E(): 3.4e-11, 26.5%
identity in 268 aa overlap. No Hp match. Contains Pfam
match to entry PF00263 Bac_GSPproteins, Bacterial type II
secretion system protein, and probable N-terminal signal
sequence"
/codon_start=1
/transl_table=11
/product="putative type II protein secretion system D
protein"
/protein_id="CAB73896.1"
/db_xref="GI:6968900"
/translation="MIRLILINILFCHYLYALDCQKRLFDISINEKLSIQESLDEL
AKYCSFSIIVKDKIAKEKLETLQNSVINHQMSLDEIFNFFIKEHDLSDYDFGKILRIS
GINTKIFKISYITISIREQSITKASVDKAPQSEYSGSFDADENMIKSMKDFQWQ
IEKEIIVLLKNSHEDYAKTPIINPNAGLIIVTQNSQLKSVKNYLOKLENRLKQVVI
IDVSIILVLSNESHSSGINMONENIGLNSOVNNSENSTIOFONGOGFYKNLGLRANLAF
DSIINFILSONGKTSVLSNPKLMALNNOQAILISVGDITINYQKESKSGTENGTTVSESY
NNYSIFVIGILLNILPEISDQKIMLRINPSLSDFKPEDNKRQKEPTIAPDTIOKKL
TSVVQVNNQPTLGLGLSHDKSNEKNSINFLSKIPILGSLFKGEVLNSKATEIVFII
TPSIVDSVNPAPSLKDLGFKHYE"
complement(185674..186936)
/gene="Cj1474c"
misc_feature
/note="Pfam match to entry PF00263 Bac_GSPproteins,
Bacterial type II secretion system protein"
complement(187013..187330)
/gene="Cj1475c"
CDS
complement(187013..187330)
/gene="Cj1475c"
/translation="Cj1475c, unknown, len: 105 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1475c"
/protein_id="CAB73897.1"
/db_xref="GI:6968901"
/translation="MCFSTFLAQDNFEYFKLIDKENRINFNTLQNPFVNPFTFKLRH
IKTITLMDKVKIYDRWYKGMIDDAIIYEINTKEIKFYDNLEIAIQINKNDKINI
NIN"
complement(187354..190914)
```

```
/gene="Cj1476c"
complement(187354..190914)
/gene="Cj1476c"
/translation="Cj1476c, probable pyruvate-flavodoxin
oxidoreductase, len: 1186 aa; highly similar
to many e.g.
NFIJ_ECOLI probable pyruvate-flavodoxin oxidoreductase
(1174 aa), fasta scores; opt: 3586 z-score: 3799.5 E(): 0,
41.7% identity in 1193 aa overlap. N-term has 28.6%
identity to HP1110, C-term has 29.7% identity to HP1111.
Contains 2x PS00198 4Fe-4S ferredoxins, iron-sulfur
binding region signature, and Pfam match to entry PF00037
fer4_4Fe-4S ferredoxins and related iron-sulfur cluster
binding domains"
/codon_start=1
/transl_table=11
/product="pyruvate-flavodoxin oxidoreductase"
/protein_id="CAB73898.1"
/db_xref="GI:6968902"
/translation="MGKIKHWGNEAAYAYATFVAGIYPTPSSPMADYDTMWA
AAGKALFGVPKIVEMQSEAGSVHGSQAGALTTTYTASQGLLLKIPNMKIAK
QLPCVHIVAAARSLAAQALISIFGDHODIYAAQIGFAMLCSSHVSQETMDLAGVAHLAA
IKGRVPFFLHFDGFTSHEIQVEMVDYAHFDRLDREALLEFRNALNPNKPTRGT
AONDDIYFOTREVSNREYDALPDVNVNMOEISKITGREYKPTTYGKHEPCVIVAM
GSVTDALREVVDYLNAGKEGVLKAVILYRPESLAYFDFVMPKSVKKLAVLDRTREPG
SLGEPLYLDKSAFAGRENAPVIGRGYGLSKQVDPAQMTAVENLKLNDKROGFTV
GLIDVTHTSISGKISLSDSTIECLFYGLGADGTGANKNSIKIIGDKTDFYQA
YFADSKKSGGYTRSHLRFKPIRSTYLVSTPHFIACVAAYLEIYDLVLAGIRKGGT
FLANSINWAEETIRQDPDAVKTLAEKEVNFYIINATKLARDIGLGNRTNTIMQSAFF
KLAKIIPYEDAQYKELAYSKSGDAIVEMNKAIDVGDAGLVKVEVDPNMKNLE
LKEKEOTNAYKGTPEVKIVKPMNAKGDLLPVSAFLGYEDGSPFHGTEYEKRGVGV
MYPRIEANCQCQCASCASVCHAVIRPFLINDEENAPRGVQHALCAKGTGKDKLS
FRQVSPDLCTGCCELCVHECTKESLVMVQEMDEFGQENADYLEKETYKDDIL
NKETTKGAQAPLQFHFHGCACGGETYITLITLFRGIMVAVNATCSSTIYGGSAF
STPYRKSVKNGHPAMGSLFEDNAEFGYGLGKMTATNTHRIETHEHNSMOEVPNALS
ALFKDWIANKONGAMSVEIKDKMIPILSONKNIKAVQDILEKQVLSKSHWIFGGDG
WAYDIGYGLDHLVLAGENVNVLVDTEVYNTGCGSSKSSRTGAVAFQAAPAKRPIQK
KDLGQIANTYGYIFVAQVNSTANTHLIKAITAAEADGPSLVICYSPCIAHGIRKGL
YSGEGELATKCGYWPITFDPRLEEQGNPLTLTGKEPDWDVYEQFLMNEVRYNSL
KKANPEHAALFERNKKDAQRYQLKRIAMADYSNEVES"
complement(188611..188862)
/gene="Cj1476c"
misc_feature
/note="Pfam match to entry PF00037 fer4_4Fe-4S
ferredoxins and related iron-sulfur cluster binding
domains., score 36.60, E-value 2.4e-09"
complement(188632..188667)
/gene="Cj1476c"
misc_feature
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature"
complement(188806..188841)
/gene="Cj1476c"
misc_feature
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature"
complement(191007..191648)
/gene="Cj1477c"
gene
complement(191007..191648)
CDS
complement(191007..191648)
/gene="Cj1477c"
/translation="Cj1477c, probable hydrolase, len: 206 aa; similar
to e.g. GPH_ECOLI phosphoglycolate phosphatase (252 aa),
fasta scores; opt: 209 z-score: 257.5 E(): 5.1e-07, 27.5%
identity in 222 aa overlap. Contains Pfam match to entry
PF00702 Hydrolase, haloacid dehalogenase-like hydrolase.
Also similar to Cj1233 (27.3% identity in 205 aa overlap).
No Hp match"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="CAB73899.1"
/db_xref="GI:6968903"
/translation="MNKTIILFDLDTLIDSTAILNSFOGAKFALGLTSKNNEIKNL
IGYPLEQMRLYPMKVLNLFYREIYQIYLEQTLLPKRALELGSBIADL
GIVTTKGKTPILLDYLGVKKRFPKTLITLEDTPNPKSPSEPIITLAKLGNLKTQENAY
MIGDTIIDQAAISANTPTALTCGYGNEENELKAHSMVFLNAYEAVNYIARLN"
complement(191094..191645)
misc_feature
```

/db_xref="GI:6968893"
/translation="MLKORLDEVNAILAKLIALTEEDIENIKVAKHESVTPSPVEEKNK
LIAEFTAKOLDVALVENSSTKGLSELLDEDKOKLDLKKNLQNLHLSKNKEYAK
FVLIVDFDLGLVNMKFDINDGNNAVGDKKTPESIFKINV"
178337..180163
/gene="flgK"
178337..180163
/gene="flgK"
/notes="Cjl466, flgK, possible flagellar hook-associated
protein, len: 608 aa; some similarity to e.g. FLGK_ECOLI
flagellar hook-associated protein 1 (546 aa), fasta
scores: opt: 261 z-score: 274.8 E(): 5.5e-08, 22.9%
identity in 612 aa overlap, 36.4% identity to Hp119.
Contains pfam match to entry PF00460 flg_bb_rod, Flagella
basal body rod proteins"
/codon_start=1
/transl_table=11
/product="putative flagellar hook-associated protein"
/protein_id="CAB73889.1"
/db_xref="GI:6968894"
/translation="MGIFGTLTGVTGLKASEVQIATGNNISNANATFYTRRVVQT
TNGITITGVQVGTAVESIVRLHDEISYIKLKGASTQLEYTRYWASTILOEIAQRPP
DLQNTGLLODENYKAWDNFASPNENATKIALVKASQTLTESVNNFATLDKIQKK
VNDIKNTVDENIRIGEETATINKIYQGEALPTEHANELRDRDELTLSLVSAY
ASKNEINQDNRLDTITDPGHQYNLSIEGSIVDGINFHLKLDYDDKKNSYIYVET
PDKVRDLAKISGGOLGALDLGRNYSKSEGYEDGIIQGYMDSLDPAKTMINET
NMLYASAKSVTSDYLSGLKGDIPLVNVDRITOPGSFDIIVDDKGDKLTKTITID
VNTTMDIRQINANDNDNKNNDVDHINASFYDAKTGDGLGFINAKSGFKVA
IEDQNFAGAFSGIFFSGTGDASDMKVDKSIDLNDPSTVRASGSDVSGNDMANKIIQ
LOKKYNFNVEDGTIDNLTMEEYRKLTGKIASDGNNNVNSNETLYNSVYSEYQS
KSGVNTNEELAAIIQYSSYGAAAKIVSTVDQMLDTLLGLKS"
178355..178447
/gene="flgK"
/note="Pfam match to entry PF00460 flg_bb_rod, Flagella
basal body rod proteins, score 26.40, E-value 4.4e-05"
180180..180944
/gene="Cjl1467"
180180..180944
/gene="Cjl1467"
/notes="Cjl1467, unknown, len: 254 aa; some similarity to
TR:O51480 (EMBL:AE001154) Borrelia burgdorferi BB0530 (217
aa), fasta scores: opt: 331 z-score: 404.6 E(): 3.2e-15,
32.8% identity in 189 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cjl1467"
/protein_id="CAB73890.1"
/db_xref="GI:6968895"
/translation="MKNFTHLKAKLDFLANOKNTNHSLFETPPDLPLOIAKIHDEFIAL
ICALFAYGNAKIVNFKLKLDFSLNLOEKQIKKELKNLYRQNEKDIOEIFITLSR
LKNEISUYELFYQAYOERENTDAILAFMQIKTLNSYSYGYDFFGKIWNTPISR
LKRYNMYLRMWVRKDELDLGLFTKIHTKDLPLDTHAHKISLTGLLLKRRKIYDYKSV
LEITQNLKLDANDPIKYDFALYRLGQSKRIDRFE"
180948..181712
/gene="Cjl1468"
180948..181712
/gene="Cjl1468"
/notes="Cjl1468, probable integral membrane protein, len:
254 aa; similar to hypothetical membrane proteins e.g.
YFCA_ECOLI (269 aa), fasta scores: opt: 612 z-score: 731.5
E(): 0, 43.4% identity in 242 aa overlap. No Hp ortholog"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73891.1"
/db_xref="GI:6968896"
/translation="MGFEHLNTIYFILRFVALFAGFIDSIIVGGGLITLPALIAQGI
PAHLSLATNKLQSVFGSFTATLTYFKSTPLRLAWGVFTFALGAAGSYSLVPKDEQ
LKIIILIFTLTYLALRPNLKGHESEPKNIKIFHLICGLTIGTFGDFLGGPGTGS
FWIFACVMLLGNMRRKASINTKILNFTSNIIAIAFLWQYELLWAYGLLMGVGVILGA
YLGSLVLTKNGKFIKTFLVIVGATIIKRVADYFS"
complement(181708..182885)
/gene="Cjl1470c"

CDS
complement(181708..182885)
/gene="Cjl1470c"
/note="Cjl1470c, probable type II protein secretion system
F protein pseudogene, len: 1178 bp; similar to e.g.
HOFK_ECOLI protein transport protein HOFK (400 aa) (22.7%
identity)and GSPF_PSEAE general secretion pathway protein
F (405 aa) (23.0% identity). Contains Pfam match to entry
PF00482 GSPIL_F, Bacterial type II secretion system
protein F domain. NO Hp match"
/codon_start=1
/pseudo
/transl_table=11
/product="pseudogene (type II protein secretion system F
protein)"
complement(181729..181977)
/gene="Cjl1470c"
/note="Pfam match to entry PF00482 GSPIL_F, Bacterial type
II secretion system protein F domain, score 33.50, E-value
9.2e-09"
complement(182130..182528)
/gene="Cjl1470c"
/note="Pfam match to entry PF00482 GSPIL_F, Bacterial type
II secretion system protein F domain, score 74.10, E-value
1.8e-20"
complement(182882..184441)
/gene="Cjl1471c"
complement(182882..184441)
/gene="Cjl1471c"
/notes="Cjl1471c, probable type II protein secretion system
E protein, len: 519 aa; similar to e.g. GSPF_ERWCA general
secretion pathway protein E (498 aa), fasta scores: opt:
993 z-score: 1102.0 E(): 0, 38.2% identity in 487 aa
overlap, and HOFB_ECOLI protein transport protein HOFB
(461 aa), fasta scores: opt: 925 z-score: 1027.3 E(): 0,
37.7% identity in 453 aa overlap. No Hp match. Contains
PS00017 ATP/GTP-binding site motif A (P-loop), PS00662
Bacterial type II secretion system protein E signature,
and Pfam match to entry PF00437 GSPIL_E, Bacterial type II
secretion system protein"
/codon_start=1
/transl_table=11
/product="putative type II protein secretion system E
protein"
/protein_id="CAB73893.1"
/db_xref="GI:6968897"
/translation="MESMDKIFQAVIDNEISLDEICTKFNITSWDFNKLAFNCNUH
FVDLDEDNFIYEGIPFSLLLKFKLLIKNNDGFMIRSKPCSLELLEQVKTMCCK
IDTAIDELKIAIKLQIRIQEEIKRLSIKLRLEQWENHRRDDQSCISQIFDILLHEI
LSPHSDIHTEARNNDALIRFVDGILREFAIKEDIYEAIVFHVKFLACLNLAESRK
TODGSELPDENRYDFRVSCLPIYGESVIRILKDKEDILDYKLNGLKNLEILK
KILHRPNGMILLTGPYGSGKSTLYACLNELKSIEKKIISAEDPIEKPLVQOILLN
SKVGFNSVIRAILRODDPIIMIGEIRDEESLDIALKASLTGHLLLSLHTNDALST
IDLLDMQAKSYLIASALSVIAORLVKLCPCWKQKSKKHYTEFESEFFEPKGCRC
HHSFGFRELIAECLTINEDLACAIRENQDKTILMELAKKYGTQWFEQGLKKAQEL
TSDLELLVVVR"
complement(183164..183949)
/gene="Cjl1471c"
/note="Pfam match to entry PF00437 GSPIL_E, Bacterial type
II secretion system protein, score 428.90, E-value
6.9e-127"
complement(183353..183397)
/gene="Cjl1471c"
/note="PS00662 Bacterial type II secretion system protein
E signature"
complement(183551..183574)
/gene="Cjl1471c"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(184431..185018)
/gene="Cjl1472c"
complement(184431..185018)
/gene="Cjl1472c"
/notes="Cjl1472c, probable membrane protein, len: 195 aa; no
Hp match"

of the E. coli dut gene. H.p. has an E. coli-like dut gene"

/codon_start=1
/transl_table=11
/product="putative dutPase"
/protein_id="CAB73874.1"
/db_xref="GI:6968879"

/translation="MTNIEILNNMLKQKLNDETNGLNWNGYTKRGKLISSWRRCIY
MECAELIDSTWTKHKNISLITWNNVRIETDIWHFILLILLPEYEDKKNKDFKAIA
TEYNAVSVDFCKEENYDIEGDIYGLINDIELIHKCSGFGNIGELLSTYFTLAIK
CGLNLEIYITTYIGKNVLNIFRONNGYKSGYKRTWNGKEDNEVLQAILQELDFDPI
YKKLEECYKKA"

gene
167285. .167998
/gene="Cj1452"
167285. .167998
/gene="Cj1452"

CDS
/note="Cj1452", possible integral membrane protein, len: 237 aa; 29.7% identity to HP0174. Contains four large hydrophobic domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73875.1"
/db_xref="GI:6968880"

/translation="MDILKLAIKDFLSKFLKALPIELFSLIIMFLGVLGFSALID
YFNLSFVGEDSFWAFYALHFVQILITIIISLFGFIVVFVAFALRITTSFLTPFI
AKETNOKYHYDNTNEVSTLKTIFEIKPIKEIGILLCTLALFPIINIFVYLYAF
YIFHLKLLMDIVSTILDKESFNHSDFSPLFEKFKSTLCFYLLSSVPLGLFLQVFF
VIFLTHLSYQRILKLKAKA"

gene
167285. .167981
/gene="Cj1453c"
complement(167981. .168946)
/gene="Cj1453c"
complement(167981. .168946)
/gene="Cj1453c"

CDS
/note="Cj1453c, unknown, len: 321 aa; similar to hypothetical proteins e.g. YACA_BACSU (486 aa), fasta scores: opt: 290 z-score: 342.2 E(): 9.7e-12, 26.8% identity in 257 aa overlap. 40.7% identity to HP0728"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1453c"
/protein_id="CAB73876.1"
/db_xref="GI:6968881"

/translation="MQIKDEILSLKRGKLLAFSYGSDSVLYFLIMQEKIDFDILVQ
INTKRNQSOLEELKALAKLKHKKIFIKHAPFQSNFEKAKARDYDFEFICILEQ
GYDHLIAHLNDQEFWFLMQLSRGAGLAILCMQCEKRSNTLLRPLLFISKDETS
SELKEDILFYHDENENKYPFRNYIKRNFSAFVSEFHQGLRSPSYLDEDRKKLYD
FENIKELQGLICPKNESLIARAVKMKGLLSTAQRKELKGDVGLGKTALAYKNEQ
AIVFEYETCOKLPKNFKECRIAKIPRLRLRAYLYNHKIDISSLSF"

gene
168936. .170255
/gene="Cj1454c"
complement(168936. .170255)
/gene="Cj1454c"

CDS
/note="Cj1454c, unknown, len: 439 aa; similar to hypothetical proteins e.g. YLIG_ECOLI (441 aa), fasta scores: opt: 671 z-score: 748.3 E(): 0, 34.2% identity in 433 aa overlap. 52.1% identity to HP0734. Also similar to Cj1006c (28.6% identity in 419 aa overlap), and Cj0458c (26.7% identity in 371 aa overlap). Contains PS01278 Uncharacterized protein family UPF0004 signature, and Pfam match to entry PF00919 UPF0004, Uncharacterized protein family UPF0004"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1454c"
/protein_id="CAB73877.1"
/db_xref="GI:6968882"

/translation="MSKLYILMSLGCNKNLVDSIMLGHLSAYELCDPEPSKADVLIVNT
CGFIDSAKRESINAILDHEORKDLSLVVTGCMORYREELMKELPEVDLFTGVGDY
ERIDEMLTAKTNLFNSYVLOSNSKRITGNSHAFIKIAEGCNKQCSFCAIPSPKG
KLKRSREISIIAELKDLVARGYKDFEIAQDTSYSLFDKGKGLRLRIDEVEKIKGI
RAARILYPTSASEALIKRIASEIFINFDMPQLQHIRSDNMLKIMKRGANSTRLEKEM
LNLMKAPNSFLRTGIVGHGSEADFEELCFVDFGDFIRSVFAYSKEDTAAPD
MEQVPFKVINKRLKIIIEKIVDIEVIEKSEKVGQKRLVVTGKSSEGEFFIAAKDLRW

DREIDGELINSECBGNLEMGQIYECEITLQNLDKLLAKALRKVDAN"

complement(168957. .170249)
/gene="Cj1454c"

/note="Pfam match to entry PF00919 UPF0004, Uncharacterized protein family UPF0004, score 454.30, E-value 1.5e-142"
complement(169776. .169838)
/gene="Cj1454c"

/note="PS01278 Uncharacterized protein family UPF0004 signature"
170329. .171426
/gene="prfB"
170329. .171426
/gene="prfB"

/note="Cj1455, prfB, probable peptide chain release factor 2, len: 365 aa; similar to e.g. RF2_BACSU peptide chain release factor 2 (366 aa), fasta scores: opt: 1138 z-score: 1344.5 E(): 0, 49.2% identity in 364 aa overlap. 67.9% identity to HP0171. Also similar to Cj1612 prfA (37.0% identity in 354 aa overlap). Contains PS00745 Prokaryotic-type class I peptide chain release factors signature, and Pfam match to entry PF00472 RF-1, Prokaryotic-type class I peptide chain release factors"
/codon_start=1
/transl_table=11
/product="peptide chain release factor 2"
/protein_id="CAB73878.1"
/db_xref="GI:6968883"

/translation="MDNYEFSELLTKLNKVGNTASIIKPNITQTRIKETIEELNSPS
FWSDVQOAGIIGKEKTKITNLLKNYENAFNALDASELFDLANSENDTETLEALFNDA
KPLEDTTSEIEISMLSGENDGKNAIVSIHPGAGGTESNDWASILYRMVLRFCEREGF
KVETLDFQEGEAGLDVSPFLVGNAYGYLKAENGIRHLVRTSPDPSAGRRHTSFSS
VMVSPELDDIEIEIEKDIRIDYIRASGAGGQHVNKTESAVRITHPTPTGIYVQCQND
RSQHKNKATAFKMLKSLYELELMQODSANTSEKSEIGWGHQIRSVLSPYQQVKRN
RSGEAFSQVDNILDGDIKMKIEGVLITAKAE"

gene
170734. .171333
/gene="prfB"

/note="Pfam match to entry PF00472 RF-1, Prokaryotic-type class I peptide chain release factors, score 368.30, E-value 1.3e-118"
171058. .171108
/gene="prfB"

/note="PS00745 Prokaryotic-type class I peptide chain release factors signature"
complement(171584. .171898)
/gene="Cj1456c"
complement(171584. .171898)
/gene="Cj1456c"

/note="Cj1456c, probable periplasmic protein, len: 104 aa; no Hp match. Contains possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73879.1"
/db_xref="GI:6968884"

/translation="MKKILTSALGAMTLIVDCGDLVELKNGNDQAMDLGNQIE
IMKKYNYKLYDYTSVRTISDINNPECKEMKARDLEIDKTPYSNYVFLNPNDKKRI
KL"

complement(172020. .173138)
/gene="Cj1457c"
complement(172020. .173138)
/gene="Cj1457c"

/note="Cj1457c, unknown, len: 372 aa; similar to hypothetical proteins e.g. YGBO_ECOLI (349 aa), fasta scores: opt: 345 z-score: 393.7 E(): 1.3e-14, 29.1% identity in 368 aa overlap. 44.5% identity to HP0926. Contains Pfam match to entry PF01142 UPF0024, Uncharacterized protein family UPF0024"
/codon_start=1
/transl_table=11
/product="ypothetical protein Cj1457c"
/protein_id="CAB73880.1"
/db_xref="GI:6968885"

```
/codon_start=1
/transl_table=11
/product="putative capsule polysaccharide export system
periplasmic protein"
/gene="kpsM"
/db_xref="GI:6968873"
/translation="MKKILILLFSCILCGAVDVQSIIISAEQDSATTSIDQNLSNNY
DIANKENLTKIHQLVFGAHLFNGFNKNYNORVPDYKIAVGDOI SLK1WGAVEFS
QLVVDSOGNTIFPKVGAVNLLGKNSALSVIKARVNKVIKSNVYVADNAYQNV
VFTGNVAPGLYOGSSDSVIQYLDKAGGLNLEVSFRDQILRNNSVINIKDIDYF
LLKGOMDLFPRSOGDVLVGNQYAFVNGDVQRFRELSNDIKTLADLARVSGAKP
IVTNAIYSGEDHKLDVSAYNKMFQSKVLRTDEVEFHPEYYSSENTITVNGEHNG
LKTIVVGKTTLEDISKLIKANKPQSNQALQVFRKSVARTOKELINAKQLKETLALT
SGSVTAQAARAQAOKTILEFIQRAKQVEPKQIVINDPKSVNSVILEDDTINVPS
KNNLIIVOGESVLPGAFVDYDKDLRYVINLAGGVGERADTSKVLVIRSNKAKEYHS
GIDMKRGDSVLVLPKVDSENQIFESMLTQILYQIAIATNVNLN"
/gene="kpsE"
/complement(162881..163999)
/gene="kpsE"
/complement(162881..163999)
/gene="kpsB"
/translation="Cj1445c, kpsE, probable capsule polysaccharide
export system inner membrane protein, len: 372 aa; similar
to e.g. KSTL_ECOLI capsule polysaccharide export
inner-membrane protein (382 aa), fasta scores; opt: 752
z-score: 755.4 E(): 0, 34.2% identity in 351 aa overlap.
No Hp match"
/codon_start=1
/transl_table=11
/product="putative capsule polysaccharide export system
inner membrane protein"
/gene="kpsB"
/db_xref="GI:6968874"
/translation="MEENKTLLEKIKDLSIFDSFKIVWFIMFVIVYVTLIAPRYVS
TMILDVSTSGESAQTSILSLSTSTASEDLNFKGYESSDMKILDEKIKKLN
YQEHIDLPKINDSSIESYLTQYQARVHTDETTLLKLVVEGFTPKSAHLIAQA
IMQSEKFINESHKAREQMAFENEQVKYKERYQAQNDLIAFONKYGVDFPLKQA
ETKASLVAEANLAQAEKALLTLQSYMNDAAPEVALKAEIAEIKQLLREKSKISA
NNSQKLNDAAKQPDLTIEATFAQYAEALKAAYESARIEALRKIQOLVLIQSDIP
ESAKYPERIVDILTAFTVLSLIFGIVKFIKMIIEHKY"
/complement(163999..164661)
/gene="kpsT"
/complement(163999..164661)
/gene="kpsT"
/translation="Cj1447c, kpsT, probable capsule polysaccharide
export ATP-binding protein, len: 220 aa; identical to
TR:069286 (EMBL:AJ000856) C. jejuni ABC transporter
protein (220 aa), and highly similar to e.g. KSTL_ECOLI
polysialic acid transport ATP-binding protein (219 aa),
fasta scores; opt: 810 z-score: 945.3 E(): 0, 58.7%
identity in 218 aa overlap. No Hp ortholog. Contains
PS00017 ATP/GTP-binding site motif A (P-loop), PS00211 ABC
transporters family signature, and Pfam match to entry
PF00005 ABC_tran, ABC transporters"
/codon_start=1
/transl_table=11
/product="putative capsule polysaccharide export
ATP-binding protein"
/gene="kpsT"
/db_xref="GI:6968875"
/translation="MKLNLNLTYSPLFSGRRHYVFNFTFEPPENCISGLMGYNGAG
KSTLMRLLSGAELPDGSKIITNKKLSWPLGLNGAFTARDNAFVARVYKGYKKE
LQEKVFEVDFAEGLKFFDEPKMTYSSGMSARTAFGLSMAFDXYLIDPAGAVGPA
FREKSVKLYERLSKSKVINVSINVAIEKWCDCIKIYMKNGQITVYDDVDEGIAVYQG
KV"
/complement(164059..164568)
/gene="kpsT"
/notes="Pfam match to entry PF00005 ABC_tran, ABC
transporters, score 86.30, E-value 6.3e-22"
/complement(164239..164283)
/gene="kpsT"
/notes="PS00211 ABC transporters family signature"
/complement(164524..164547)
/gene="kpsT"
```

```
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/complement(164658..165440)
/gene="kpsM"
/complement(164658..165440)
/gene="kpsM"
/notes="Cj1448c, kpsM, probable capsule polysaccharide
export system inner membrane protein, len: 260 aa;
identical to TR:069285 (EMBL:AJ000856) C. jejuni ABC
transporter protein (260 aa), and highly similar to e.g.
KPMI_ECOLI polysialic acid transport protein kpsM (258
aa), fasta scores; opt: 488 z-score: 558.3 E(): 8.9e-24,
34.3% identity in 254 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="putative capsule polysaccharide export system
inner membrane protein"
/gene="kpsM"
/db_xref="GI:6968876"
/translation="MLNVIYALFRELKTRFGKNRYLGYINWVGPEMSIVLLVPIIGT
IVREYHHQVMEGISIFNLISGIMPFRSIVTQLMNGTOANALFAYKVPKPIHV
FIATLLEFCIYFVIFIIILFFAGWFFRLDVPFVHLLGVLCIFLICSAPALGICFA
LIWHFPEPLRTLAYFSIVFYWTSGIIFPTWLTPTPLDIDIFYNPLLIHIFELLRFNFF
ENYPLQDEYSYFYAIFWILLVLFICLFIYYVNRQALTAVKKE"
/complement(165512..165919)
/gene="Cj1449c"
/complement(165512..165919)
/gene="Cj1449c"
/notes="Cj1449c, unknown, len: 135 aa; similar to
YHBP_ECOLI hypothetical protein (147 aa), fasta scores;
opt: 236 z-score: 315.0 E(): 3.2e-10, 30.5% identity in
131 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1449c"
/gene="Cj1449c"
/db_xref="GI:6968877"
/translation="MDERILLEFKNEOLLSWAMIDEXGYVTASAFYAFDEKNLAFIIA
SHEDTKHIRLASSENSIALNIAKESKIAFLKGQVAKAEFRMASKEQMKIYFSKPPFAK
FDSAKIYALELFWLKFTNNALGSKKLEFYK"
/165994..166557
/gene="Cj1450"
/165994..166557
/gene="Cj1450"
/notes="Cj1450, probable ATP/GTP-binding protein, len: 187
aa; no Hp match. Contains PS00017 ATP/GTP-binding site
motif A (P-loop), and probable coiled-coil domain at
C-terminus (aa 150-175)"
/codon_start=1
/transl_table=11
/product="putative ATP/GTP-binding protein"
/gene="Cj1450"
/db_xref="GI:6968878"
/translation="MQIDASKNOSFSDMTYTKSKGHLALSMYDNOSVSYANNEEGKTL
NLKROYGFSFTFEGSKLTQNDLDEIKNAMEVEPMTKDFLANSKVGELKPELIESAM
QMANVLPFPNDENHONAMNFTNKLSDLLKQKQTDKDDKINASMLEDSKLLDEVLEQ
MKKOLEKQEKAKENQDKTDDSLNLYA"
/166099..166122
/gene="Cj1450"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
/166594..167283
/gene="dut"
/166594..167283
/gene="dut"
/EC_number="3.6.1.23"
/notes="Cj1451, dut, possible dUTPase, len: 229 aa; similar
to two enzymes of eukaryotic origin e.g. TR:O15826
(EMBL:L76742) Leishmania major characterised dUTPase (EC
3.6.1.23) (dUTP pyrophosphatase) (268 aa), fasta scores;
opt: 311 z-score: 382.3 E(): 5.7e-14, 33.6% identity in
226 aa overlap, and TR:O15923 (EMBL:U93211) Trypanosoma
cruzi deoxyuridine triphosphatase (283 aa), fasta scores;
opt: 198 z-score: 246.6 E(): 2.1e-06, 32.9% identity in
228 aa overlap. No Hp match. C.j. does not have a homolog
```

misc_feature

gene

CDS

/db_xref="GI:6968870"
/translation="MKIVIVGIGYVGLANAILFSGKNNENEVLLIDENKIQSINNHH
SPTKDLIEKFFVQNTSKLHATNIKEAVFNADFAVIATPTDYDEOLNFPDFRSIENY
LKDKNINSKINVIKSTVPIGYTKTKOKFNMSNIVESPERTRESALYDSLYPSRI
IIGDKSVLGTIGDLFLKNIIEKKANVDIFYMDSDEAESVKLFNTYLA MRVGFNFNEVDS
YARKHNLNADIIGISADDRIGKYNPNPSFGYGGYCLPKDKTKOLIANFYINPSLFIK
AVIVETNEIRKFTQILIEKKPNLGIYLRIMKONSDFNRSNIIDIIKYLOYSNSI
ELIIVEPLVREKKFLNIKVENDFNFGAKVDLIIANRFDKLEIKDKVFSADVFYTD
I"
complement(157460) .158632)
/gene="kfid"
/note="Pfam match to entry PF00984 UDPG_MGDP_dh,
UDP-glucose/GDP-mannose dehydrogenase family, score
247.50, E-value 1.9e-70"
complement(158629) .160263)
/gene="Cj1442c"
complement(158629) .160263)
/gene="Cj1442c"
/note="Cj1442c, unknown, len: 544 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1442c"
/protein_id="CAB73866.1"
/db_xref="GI:6968871"
/translation="MPLLSVIIIPFGLSKERPYTEERVLOKAKPFQSDENIEFIFVEGY
SKDHLEKNFIOANHIIYLKMDQKAFSGRCRNLGASVHNSVLLFLDVCVISTDN
FEKILKLIQIKNISONALIVLVVYLNKEANEKLGKQYDDKPMWILLOEDLFTAKNT
WYKFFAPSTSSIVINKHQLRLGNDENFIGHGYEDFLFARLKAACVSFEKMPNL
SYDARNWNFNFKPSFWSGLGYEACFYIYHIEHPNQNMAYMONKDKNHQLFY
KHLKNIKHDLKPLQVFAKGEKVLMLFKDQNYDFWDISVYGEIYIKNISDFIAKE
LKYELLDFLOEQKITKIFLDASIKLQFNGIDYCDIFYFQKGLPHSWFAKKLDLYG
KODLNLSEOKLYQVKEYFLTKDEKKVILDFLSONNDKDYLYKMLYVILNLYSFE
HGFIFQIIIDKEKMLIAQHDKSFYPLRSFYKPYLHELKSRPFSFLMKILGLEYL
GAMISHKTYRLARKLFFNPKFFEDFNFKIMKGKI"
complement(160263) .161210)
/gene="kpsf"
complement(160263) .161210)
/gene="kpsf"
/note="Cj1443c, kpsF, unknown, len: 315 aa; similar to
hypothetical proteins belonging to the kpsF/gutQ family
e.g. KSFI_ECOLI KPSF protein (317 aa), fasta scores; opt:
849 z-score: 979.3 E(): 0, 46.5% identity in 297 aa
overlap. 38.7% identity to HP1429. Contains 2x Pfam match
to entry PF00571 CBS, CBS domain"
/codon_start=1
/transl_table=11
/product="Kpsf protein"
/protein_id="CAB73867.1"
/db_xref="GI:6968872"
/translation="MNTLEIAKEVFEKAQAAILDLATNLNENFNQAVNLMLNTKGRCI
VSGMKSGHIGAKIAATLASTGTPSFFIHPGALHGDGLMGLTSEDVLIASNSGETEE
ILKIIPAIAKREIPLIVMCGKNSILYKOGDIFLNIAVEKEACPQLQAPMSSTATLY
MGDALAAKLVKNRNPEDLFLFHPGSGSLRKLTKVKDMLVSSNLPITVHPDTEFNDL
VDVMTSGKLGVLENKGLIITDGLRALKADSKDRFDFKAKEIMSINKPNVYDA
DAMASEAGIMLKHKIKIEIIVGKEEKVVGIIQLYATGNV"
complement(160272) .160418)
/gene="kpsf"
/note="Pfam match to entry PF00571 CBS, CBS domain, score
10.70, E-value 1.3"
complement(160455) .160613)
/gene="kpsf"
/note="Pfam match to entry PF00571 CBS, CBS domain, score
41.30, E-value 2.1e-08"
complement(161221) .162879)
/gene="kpsd"
complement(161221) .162879)
/gene="kpsd"
/note="Cj1444c, kpsD, probable capsule polysaccharide
export system periplasmic protein, len: aa; similar to
e.g. KSDI_ECOLI polysialic acid transport protein KPSD
PREC (558 aa), fasta scores; opt: 1187 z-score: 1306.9
E(): 0, 36.1% identity in 560 aa overlap. Contains
probable N-terminal signal sequence. No Hp match"

/genes="Cj1438c"
/note="Pfam match to entry PF00535 Glycos_transf_2,
Glycosyl transferases, score 63.40, E-value 4.9e-15"
complement(155074) .156180)
/gene="glf"
complement(155074) .156180)
/genes="glf"
/EC_number="5.4.99.9"
/note="Cj1439c, glf, probable UDP-galactopyranose mutase,
len: aa; highly similar to e.g. TR:086897 (EMBL:AJ006986)
Streptococcus pneumoniae polysaccharide capsule
biosynthesis cluster UDP-galactopyranose mutase (EC
5.4.99.9) (369 aa), fasta scores; opt: 1546 z-score:
1775.3 E(): 0, 59.9% identity in 364 aa overlap, and
GLF_ECOLI UDP-galactopyranose mutase (EC 5.4.99.9) (367
aa), fasta scores; opt: 1545 z-score: 1774.2 E(): 0, 60.8%
identity in 362 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="UDP-galactopyranose mutase"
/protein_id="CAB73863.1"
/db_xref="GI:6968868"
/translation="MYDYLIVGSLFGSIFAYEATEKEGYTCLVREHIGGCNYTEN
IKNINHYGKAHIFRTSDQINWMQFCFENHFINSPIAYIDEIYNLFPMNMTFSK
LWGIKTPNEARKLTIEMQKQIIQHPKNLEQALISLVGYEKLIGYTEKQWGRSCK
DLPASIRRLPVRYIYDNNYFNDPYQIPKGGYTAJFDKMLKSKVILNTDFLKYDKD
FKNAKKIVFTGICIDAYDYRYGALEYRSILKPEHKILNLDNFGQVAVVNTYDKIEIPT
RIIEHKHFEFGNDTIVISEEYPLEWIKGIEPYYPINDEKNQALYKYLQALAHESNV
YFGRLGEYRYDDQMVRSALLFCNKLNLKNGG"
complement(156218) .157441)
/gene="Cj1440c"
complement(156218) .157441)
/gene="Cj1440c"
/note="Cj1440c, probable sugar transferase, len: 407 aa;
similar in N-terminus to e.g. TR:007340 (EMBL:007340)
Streptococcus pneumoniae SS-1,4-galactosyltransferase (318
aa), fasta scores; opt: 280 z-score: 326.2 E(): 7.5e-11,
26.7% identity in 374 aa overlap. No Hp ortholog. Contains
Pfam match to entry PF00535 Glycos_transf_2, Glycosyl
transferases"
/codon_start=1
/transl_table=11
/product="putative sugar transferase"
/protein_id="CAB73864.1"
/db_xref="GI:6968869"
/translation="MKTGVGVPIYVYVYKYLRECLDSVVNQYKKNLQVVLVNDGSTDE
NSLNAKEITLADERILFDKENGSGSTARNVGIEFFSKEYDFKNITQELKENSILVEF
KLDNEDPNYIYKYSNFFKNKDLNFKAPDIDYIIIFLSDDDYWEINLCEECVPK
MDGEVYWFDSGSIFEEGFKQWSSLLKLYDLHEGVKSKVMLEYISINKKIYNFYFTW
SGMIDFYLNKILKFDIDYIIHQDHHGMLLFAKCKYIVIFPSSMHTYIRISNTINL
SDSELOIHIPRYMOSIYDSFNINIKEAKEYSSMSMFYVLKNCFNDKILMENRYNAYL
LEKAPFKFICERCLRIFLLRDPCCKIQIKKILKQYKNRYVHLTNFTNGATYIFKIF
KNIINIKYIRLYFFR"
complement(156995) .157429)
/genes="Cj1440c"
/note="Pfam match to entry PF00535 Glycos_transf_2,
Glycosyl transferases, score 24.30, E-value 0.00048"
complement(157451) .158632)
/gene="kfid"
complement(157451) .158632)
/gene="kfid"
/EC_number="1.1.1.22"
/note="Cj1441c, kfid, probable UDP-glucose
6-dehydrogenase, len: 393 aa; similar to many e.g.
UG6_ECOLI UDP-glucose 6-dehydrogenase (EC 1.1.1.22) (392
aa), fasta scores; opt: 1386 z-score: 1511.4 E(): 0, 52.3%
identity in 392 aa overlap. No Hp match. Contains Pfam
match to entry PF00984 UDPG_MGDP_dh,
UDP-glucose/GDP-mannose dehydrogenase family"
/codon_start=1
/transl_table=11
/product="putative UDP-glucose 6-dehydrogenase"
/protein_id="CAB73865.1"

misc_feature
gene
CDS
misc_feature
misc_feature
gene
CDS

```

/note="Cj1434c, probable sugar transferase, len: 445 aa;
similar in N-terminus to e.g. TR:007340 (EMBL:007340)
Streptococcus pneumoniae SS-1,4-galactosyltransferase (318
aa), fasta scores; opt: 344 z-score: 405.7 E(): 2.8e-15,
27.3% identity in 326 aa overlap. No Hp ortholog. Highly
similar to Cj1438c (72.4% identity in 456 aa overlap).
Contains Pfam match to entry PF00535 Glycos_transf_2,
Glycosyl transferases"
/codon_start=1
/transl_table=11
/product="putative sugar transferase"
/protein_id="CAB73858.1"
/db_xref="GI:6968863"
/translation="MMNVTPEKSIIVVPSLNSISVIRECIDSILNOTLKDIEILCIDA
NSTDGTLEVLNVEKKDKRLRLVIIISDKKSGYQNMNLGKEAKGEYLGIVESDDYIKTN
MYERLYEIAKNDCVEVKGDFYIFAYGKTEYVNLVNSCEDIYNNKWNKDIRIFIG
SDGINPIGIYRLDLRTNQIKLNETPGASYODNGLWFQIFALAKSYIFNEAFYMLRR
DNPSSVKKEKVCACEEYDFIRDLKKHPDLEKTAPICALHFRFGMYFTLERIDE
RYKLDLFRFSQDFRKILKDELDELNFNGINMQRINKIIENPVYIYVFSRGARLQ
NOLVRLGKVVVEAKSFNKIIKLPFLMKICLHNFHVKVRSIVQRPDLKLLPLEC
YLDYHEALVIVKEHLSYKFGKILLISFKGWYKGKIFILPFLMKRYKYEKNMI"
149216..149251
complement(149218..149712)
/gene="Cj1434c"
/note="Pfam match to entry PF00535 Glycos_transf_2,
Glycosyl transferases, score 70.80, E-value 2.9e-17"
complement(149791..150426)
/gene="Cj1435c"
complement(149791..150426)
/gene="Cj1435c", unknown, len: 211 aa; similar to TR:Q58989
(EMBL:U67599) Methanococcus jannaschii hypothetical
protein MJ1594 (211 aa), fasta scores; opt: 177 z-score:
228.6 E(): 2.1e-05, 23.5% identity in 196 aa overlap. No
Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1435c"
/protein_id="CAB73859.1"
/db_xref="GI:6968864"
/translation="MSREIALDFCECTLNFQTLRYLPAGSKNNIYTSQKNLARR
ERFORENLPYREWLIDLDVAEEIAQEFYTDVMANLNQNMVMDRLFWHQDEGHTI
VIVSGGTLTYKAPRIYNIENIVAVDLIYKNKLTGNIDGITHPMQERKLYKLAQKN
LKQFDLKNAYSDCVSDIPLLSLVGNPNVTECKGLQWARILGFENILLYK"
complement(150430..151602)
/gene="Cj1436c"
complement(150430..151602)
/gene="Cj1436c", probable aminotransferase, len: 390 aa;
weakly similar to e.g. HIS8_BACSU histidinol-phosphate
aminotransferase (360 aa), fasta scores; opt: 304 z-score:
351.9 E(): 2.8e-12, 25.5% identity in 318 aa overlap. No
Hp match. Also similar to Cj0317 (hisC) (25.7% identity in
304 aa overlap), and Cj1437c (23.1% identity in 347 aa
overlap). Contains PS00105 Aminotransferases class-I
pyridoxal-phosphate attachment site"
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="CAB73860.1"
/db_xref="GI:6968865"
/translation="MLIKLNDYKNITQIKDLKNASGSHSPSTFWAEQIPELNKI
DSCFLSNPYATALFLRYLKEELIDGKLRSLVPEYPSQNSIIAKTVADFIDGPKNVF
IGNAEILQAVRHNFGVKLIIVNPITFYSSYEFKSETNVVYQUSKEDYNNIEH
YLNFAKNENPDSVVLINPNPDGGYINYEKLRIYLSLKYKNIIDESTIHFAYENK
DYNGINTEYLFKEFHNTIIIKSMKDFGVAGIRIGAIMSEDKIYLLKNGLWNSG
LSYFLLVYRKNFDFDEYDKVRREYIOETOTFRKLSGIKQEFKVPSPMANFALVELLD
GSSSTDPAVKMLIKYGIYMTCDNKIGLEGEFIRIASRTLEENDWLKSCDVPKE"
complement(150844..150885)
/gene="Cj1436c"
/note="PS00105 Aminotransferases class-I
pyridoxal-phosphate attachment site"
complement(151652..152755)

```

```

CDS
/gene="Cj1437c"
complement(151652..152755)
/note="Cj1437c, probable aminotransferase, len: 367 aa;
similar to e.g. HIS8_BACSU histidinol-phosphate
aminotransferase (360 aa), fasta scores; opt: 408
z-score: 478.1 E(): 2.6e-19, 28.6% identity in 360 aa
overlap. No Hp match. Also similar to Cj0317 (hisC) (29.9%
identity in 344 aa overlap), and Cj1436c (23.3% identity
in 347 aa overlap). Contains Pfam match to entry PF00222
aminotran_2, Aminotransferases class-II, and PS00599
Aminotransferases class-II pyridoxal-phosphate attachment
site"
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="CAB73861.1"
/db_xref="GI:6968866"
/translation="MQANKNIQKLTPLYSIPHKIWNSSQSNILKDWNEATIPSPYIV
IESKAEVLNGLNWYIPNKNLYLIDDKIAEYTKQINSFVELFGSSAHECIIDVFL
DKDKIGIVSYTDNFRSRANGVETISFTLDNFDLDSLEYFHEKRKILLYLC
PNNPTKSYNIQIKSLIINNPNVMTIDEAYEFTSQSVDLVEOCNNLIITRTFS
KAFALASFRIGYLIISHPENIENSINKLRNPKSPVMLSOIAANAALDQLQMYRDYDVEVS
CARMEFVKFLNTLTGGGFIENSDVANEVLQINENISLFVGFLEKEGIFIRNYSHLIS
KNCRISIGTRNMSYVAEKIQEPAKKQGGFHLV"
complement(151742..152524)
/gene="Cj1437c"
/note="Pfam match to entry PF00222 aminotran_2,
Aminotransferases class-II, score 27.40, E-value 5.6e-09"
151871..151879
/note="C(9)"
complement(152081..152110)
/gene="Cj1437c"
/note="PS00599 Aminotransferases class-II
pyridoxal-phosphate attachment site"
complement(152742..155072)
/gene="Cj1438c"
complement(152742..155072)
/gene="Cj1438c", probable sugar transferase, len: 776 aa;
similar in N-terminus to e.g. TR:007340 (EMBL:007340)
Streptococcus pneumoniae SS-1,4-galactosyltransferase (318
aa), fasta scores; opt: 325 z-score: 375.2 E(): 1.4e-13,
26.9% identity in 316 aa overlap. Similar in C-terminus to
TR:O31348 (EMBL:X11138) Bacillus cereus hypothetical
protein (ORF2) (300 aa), fasta scores; opt: 355 z-score:
409.9 E(): 1.7e-15, 31.2% identity in 340 aa overlap.
27.7% identity in 260 aa overlap to Hp0102. Also highly
similar to Cj1434c (72.9% identity in 457 aa overlap) in
N-term, and similar to Cj1421c (34.6% identity in 127 aa
overlap) and Cj1422c (23.5% identity in 452 aa overlap) in
C-term. Contains Pfam match to entry PF00535
Glycos_transf_2, Glycosyl transferases"
/codon_start=1
/transl_table=11
/product="putative sugar transferase"
/protein_id="CAB73862.1"
/db_xref="GI:6968867"
/translation="MMNVTPEKSIIVVPSLNSISVIRECIDSILNOTLKDIEILCIDA
NSTDGTLEVLNVEKKDKRLRLVIIISDKKSGYQNMNLGKEAKGEYLGIVESDDYIKTN
MYERLYEIAKNDCVEVKGDFYILESNKGYKSTIPDIDLYNQIISKTPHNPFGOS
INPIGITYRLDLRTNQIKLNETPGASYODNGLWFQIFALAKSYIFNEAFYMLRRDP
NSSVKKEKVCACEEYDFIRDLKKHPDLEKTAPICALHFRFGMYFTLERIDERYK
YDFASERIKWQSLRIGKLLIDLKNPVQILKFPFLFLEIKQKFEOKIYKTKIKYVP
NLQPLPEEYSDYEQALGKTKHLSYILGKSFINNPILEFIKIKIKYQYKDKDSSSKK
NIALSDNHEYSWDKIDILNKSILFNNDLQDKIFETNCKYLPKTYGTYKNIYD
INFNELPNSVLKTNHDCGGYIVVENKQFLRDTVFNAMKLLKHLREPNYVSFRF
WHYKDIPEYRFAPELLIGENKPADTYKPHIFDKENLSNNFIQVTTDRFNYORAMFD
LSWNLAPFENMYDNKNVTMIPKPNLLDSMINTSLILAKFPDYVRVDLYQDFDKIYIG
ELTFTHGAAGEKVIPEWDKKLGLDLWRLKRLDNASK"
complement(154578..155045)
misc_feature

```

gene /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(140931. .141857)
/gene="Cj1429c"
CDS complement(140931. .141857)
/gene="Cj1429c"
/note="Cj1429c, unknown, len: 308 aa; no Hp match.
Contains C(9-10) polymorphic region at aa 98. C10
(consensus) gives this ORF. C9 would give truncation after
+1 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1429c"
/protein_id="CAB73853.1"
/db_xref="GI:6968858"
/translation="MEKLKELNNIKLSNOLSVQILMSNIISNLDDKKDKREILLLL
LQARDNRYIRINNEOCYONILYNLRIPLCLDRLRIGNGGGGYVMYNGSGDM
SDINAKLSYSGSDSPFMLEMAQRGKVLVEYDASIEKCPYRHENIVFHKKFGNGVNN
ENTITIVQALKDNNDENPKNICODIENCEWDMLENIDISILNKYSQVIFFEHGCN
PEQDQVEKRISLKKLNEYFVPMPTFHNHGKIFYSQGLFFSTTLEVSYLRLNLI
DINKYRDVAGGFKNLDPFWISNPEIPFRGY"
misc_difference 141557. .141566
/note="polymorphism C(9-10)"
complement(141861. .142406)
/gene="Cj1430c"
CDS complement(141861. .142406)
/gene="Cj1430c"
/note="Cj1430c, probable nucleotide-sugar
epimerase/dehydratase, len: 181 aa; similar to e.g.
RFB_C_ECOLI DTDp-4-dehydrorhamnose 3,5-epimerase (185 aa),
fasta scores: opt: 303 z-score: 386.2 E(): 3.4e-14, 33.9%
identity in 165 aa overlap, and TR:Q46770 (EMBL:U23775) E.
coli im1c DTDp-6-deoxy-1-lyxo-4-hexulose reductase (183
aa), fasta scores: opt: 306 z-score: 390.0 E(): 2.1e-14,
33.5% identity in 164 aa overlap. No Hp match. Contains
Pfam match to entry PF00908 dTDP_sugar_isom,
dTDP-4-dehydrorhamnose 3,5-epimerase"
/codon_start=1
/transl_table=11
/product="putative nucleotide-sugar epimerase/dehydratase"
/protein_id="CAB73854.1"
/db_xref="GI:6968859"
/translation="MAIEFDIQESKILKGVIITPNKFRDLRGEIWTAFTEYLSKLV
PDGIEKFKDKFINSHFNVLRGIDGVTKLVTCVYGEVHVVDCKRDSPTYLKWKEK
FIISYKNQOILLPPNMGNSHVSSKEAVYYKLAYEGEYMDAPDQFTYAWNDERIGI
DWPTNPILSDRDILATKNGK"
misc_feature complement(141873. .142403)
/gene="Cj1430c"
/note="Pfam match to entry PF00908 dTDP_sugar_isom,
dTDP-4-dehydrorhamnose 3,5-epimerase, score 55.00, E-value
4.1e-13"
/complement(142406. .144154)
/gene="Cj1431c"
CDS complement(142406. .144154)
/gene="Cj1431c"
/note="Cj1431c, unknown, len: 582 aa; no Hp match. Similar
in part to N-termini of Cj1421c (38.6% identity in 114 aa
overlap), and Cj1422c (38.6% identity in 114 aa overlap).
Contains PS00455 Putative AMP-binding domain signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1431c"
/protein_id="CAB73855.1"
/db_xref="GI:6968860"
/translation="MDSNPLIIAGRDDGFGFRMRALLNALYISKSGFKFGFVWRDI
NNIQNLGKVLIPWANLPTREYFDQDEIKSYRQDIEFAVETPVLMWSLYRQSIKNI
LKPYKEGWGYSTQGLDSYEYTDVDEGYRTELVELSCWQIDFSSHVKKIFKASIKFI
LDIGKEVALHIRTGEYIHDFEYRNLHYCRYKIFPPYFALEIALKIKKGRHVFPGD
DLNLQALKEYGSFNKQAOENIFSIDIIAFEOLDNGYDRLLFELVLMKSEYIFGSG
TTGFSFCASWIKNIKIFINFDHLSLEQVEYLKLYDIDENIDDLVSCNYPFLFLSE
QLNLGDKRLYLSLRDYSGLNSEYFYINLLQNERFKKADDRLEQVICKRKKFK
FDLLGQGNQPTFPFYDIWMYFKDQSYNIFYVACRIFSEFNIFSRVNTYIPNFH
PIFDQFKMFIKDPKSDQETGAVKKIRNHLAYKLGVAAIKNSKSLGWYIRMPYVLS
YTRDMHKSQNMKDKKSIISLEYSDYESALKKEGKGVYKLGQIIKAHKNWHKGGYIM

misc_feature LWFEVKKLKKFKKKGK"
complement(143309. .143344)
/gene="Cj1431c"
/note="PS00455 Putative AMP-binding domain signature"
complement(144157. .147252)
/gene="Cj1432c"
CDS complement(144157. .147252)
/gene="Cj1432c"
/note="Cj1432c, possible sugar transferase, len: 1031 aa;
weakly similar to e.g. TR:Q48232 (EMBL:X78559) Haemophilus
influenzae orf3 in DNA for serotype B capsulation locus
(1215 aa), fasta scores: opt: 361 z-score: 398.4 E():
7.2e-15, 22.8% identity in 1029 aa overlap, and TR:O84910
(EMBL:AF010183) glycosyltransferase WBPZ (381 aa), fasta
scores: opt: 161 z-score: 182.7 E(): 0.0074, 23.1%
identity in 251 aa overlap. Contains Pfam match to entry
PF00534 Glycos_transf_1, Glycosyl transferases group 1.
Also similar to Cj1434c (23.1% identity in 424 aa
overlap)"
/codon_start=1
/transl_table=11
/product="putative sugar transferase"
/protein_id="CAB73856.1"
/db_xref="GI:6968861"
/translation="MIKKILLITPELEYTGALNSFRICEVLNNKYAVDITWYNEGP
YISEFDKLVGVVEVISEDIDSKVHERISKYSLVITANTIVVYKVELIQNLTPVVMY
IREAENLPDFFWKPERKLALEKAKLYVYSEYAKDFLIHNNKNVVELHNVVDVRYE
KHDDFLKOJSDKSLKFLAIGTIEKRKYDVLQAFIDLPVDIRQCELHPAGFEWGA
KDFPFLSLAKFPNIFYHGLDRKKHLSLIFQCNVMVWVPSRDECSLVALEGAMN
SKPLILNIGAKYIILDENSGWLVTGSDVSLNAFIQAYKKNKLDAMGANGNNYL
QSTYIEYKNIILKVRDEICKQYLYRINQENYVLFSEFIDTILSRNLAKPSAVFL
INKKRMNDPFLNLVKNDFRIVEYQVYRVNCKYEDTFDETYNLQCNFSL
FOOKEMLKLEINTEKETLYPIKKNIELVEELIKNEKRVLLISDMYFSSSIIETLNLK
FSPFNPIIYMSSEFRKNSGLKALINLEKVPDKWIGHGDNWVGQVGLKPSNLE
ISTVYFNILLPYEEFALNNRSLMDLQKILGISKIRLENTLQDIEGVISFGAPML
LPYQWILNIALKNSIRCLYFIARDGYVLQKMTDMLIQAKINIKTYLGSRESWRE
PFRKDKLKIQLIDEDQIDKQEIFAFVCECGTGETLDYIVKRTYESNOQFKNMFPG
SLYLSKLNKTKTOSLEMLPLNENVTYGIELFVRSLOGOVLYGDKDGRVIEPDEFL
EGEALQKFRYDEYIDGVLMFMEYIVKTDNVEKIEFDNNNTYLLNLYSNYIDKFKIE
IMGNVFIINGVAKRGIYFAPRLNNKITLDQKNSFNNSVLSRCDIIRVINMDNDY
FGLIGIKRDIKSHLSYKLGKILLNINPLKWLKILVILIPILIFSHEQKKIYLTEKK
INIDLFSYKDYNAINVRNFFSQGLGELILKTSRKWNLAIVILPYKIAOLYRKKFKK
G"
complement(146257. .146526)
/gene="Cj1432c"
/note="Pfam match to entry PF00534 Glycos_transf_1,
Glycosyl transferases group 1, score 27.00, E-value
1.5e-06"
complement(147262. .148368)
/gene="Cj1433c"
complement(147262. .148368)
/gene="Cj1433c"
/note="Cj1433c, unknown, len: 368 aa; no Hp match.
Contains 9 repeats of the aa sequence KIDLNNT near the
N-terminus"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1433c"
/protein_id="CAB73857.1"
/db_xref="GI:6968862"
/translation="MORFKWFLSIIKNFKQHEKIKIDLNNTKIDLNNTKIDLNNTKI
DLNNTKIDLNNTKIDLNNTKIDLNNTKIDLNNTKIELSOLKHEHYKVLDPH
LKPITFOALEIJEIHLAESNCNCFGSHFSQIAEKEPDIEIFKQMDQLSEISKG
IVGTFRMGEPGLLNPNCIQFDITRIFPFSKAIWLVTNGILLDKQNDQFNWSCQRNK
MOIRPTKYPIKINMDLIDKDKQDYIDPLIFFNNGELEKTSWKSFLDPSGNCNHYHST
NCSMAHCYQFKDGKLFCTCTFFAHQVHFNNKNGELEVCEFFIDITYKARDYQEIILF
LSKPIPCRYCKVSOWAIEGKWRSSNKTKEVLI"
complement(148109. .148297)
/note="9 copies of 21bp repeat"
complement(148402. .149739)
/gene="Cj1434c"
CDS complement(148402. .149739)
/gene="Cj1434c"

```

CDS
/ gene="Cj1423c"
complement(135771..136436)
/ gene="Cj1423c"
note="Cj1423c, possible sugar-phosphate
nucleotidyltransferase, len: 221 aa; similar to many
predicted nucleotidyltransferases, e.g. TR:O58775
(EMBL:AP000004) Pyrococcus horikoshii hypothetical
sugar-phosphate nucleotidyl transferase PH1022 (416 aa),
fasta scores; opt: 413 z-score: 496.0 E(): 2.6e-20, 34.3%
identity in 230 aa overlap, and MPGL_YEAST
mannose-1-phosphate guanylttransferase (361 aa), fasta
scores; opt: 258 z-score: 313.4 E(): 3.9e-10, 28.9%
identity in 232 aa overlap. Contains pfam match to entry
PF00483 NTP transferase, Nucleotidyl transferase. No Hp
ortholog. Also similar to Cj1329 (35.6% identity in 216 aa
overlap)"
/codon_start=1
/transl_table=11
/product="putative sugar-phosphate nucleotidyltransferase"
/protein_id="CAB73847.1"
/db_xref="GI:6968852"
/translation="MQAIIICGGIGTRLSIKIDPKPMAPINDPKPLEFIFYELKKQ
GIKEVILAVSYKYEYIKPEDEFLGIKIRYSIEKEPLGTGGAKEKLFKVNKAYVL
NGDFFDLSKLKLNESKICALKOMNDRDYGTNVVDEQDLVISPEEVKFKQGLI
NNGIYLLTKDIFNDFALQENKFSPEEFLQENYKLLKARACIFDDYFIDIGVPEDYHFL
INN"
complement(135774..136433)
/ gene="Cj1423c"
note="Pfam match to entry PF00483 NTP_transferase,
Nucleotidyl transferase, score 57.70, E-value 2.5e-13"
complement(136424..137029)
/ gene="gmbA2"
complement(136424..137029)
/ gene="gmbA2"
note="Cj1424c, gmbA2, probable phosphoheptose isomerase,
len: 201 aa; similar to many e.g. LPCA_ECOLI
phosphoheptose isomerase (EC 5.-.-.-) (192 aa), fasta
scores; opt: 453 z-score: 548.7 E(): 3.1e-23, 39.2%
identity in 186 aa overlap. No Hp ortholog. Also similar
to Cj1149c gmbA (57.8% identity in 147 aa overlap)
EC_number=5.-.-.-"
/codon_start=1
/transl_table=11
/product="putative phosphoheptose isomerase"
/protein_id="CAB73848.1"
/db_xref="GI:6968853"
/translation="MELNSIKGHFADSLVKEQILKDNLTILIKNASLEVIKAYK
NGKNTLHAGSGSAADAQHIAGEFVSRYFDFPGIASIALTTDTISLTALGNDYGYEN
LFAQVQAQVKGDFVIGISTSGNSKNILKALFCKQKEIISIGLSAGSGAMNELCD
YCIKVPSTCPRIQEAHLIGHITICAIVEELFCGKGSCKQ"
complement(137017..138036)
/ gene="Cj1425c"
complement(137017..138036)
/ gene="Cj1425c"
note="Cj1425c, possible sugar kinase, len: 339 aa;
similar to e.g. GALI_LACHE galactokinase (388 aa), fasta
scores; opt: 186 z-score: 227.6 E(): 2.4e-05, 24.3%
identity in 371 aa overlap. Contains PS00012
Phosphopantetheine attachment site. No Hp match"
/codon_start=1
/transl_table=11
/product="putative sugar kinase"
/protein_id="CAB73849.1"
/db_xref="GI:6968854"
/translation="MKTIRTQPLRLAGGGTDINLYCDKYTYGYVLNATISYIHCT
LIRKDEGLIFDPSDINSYGESEKFLNGDKGLDFKSYINRIVKDFTKPLSFSLIH
TYSDFVSGSLGSSITLVGVYKAFAEWLNPLGVEIAKLAIVEIREDDIGVGGQD
QYAAETGGFMENKRVIVNPLRIKKNIASELEARTVLYFTNITREAKDIEEHKK
GLKGDKSLPAMHAIKODALIKMEALFRADFCTLAQILGKSWRSKIIISIVSNDLEL
RIVKLAIDNGAYSCKTSGAGAGFMFFVDPTKYNLILKALRKEQGVQVDFSTKQGV
KSWRI"
complement(137656..137703)
/ gene="Cj1425c"
misc_feature

```

```

/ note="PS00012 Phosphopantetheine attachment site"
complement(138056..138907)
/ gene="Cj1426c"
complement(138056..138907)
/ gene="Cj1426c"
note="Cj1426c, unknown, len: 283 aa; no Hp match.
Contains C(10-11) polymorphic region at aa 99. C10
(consensus) gives this ORF, C11 gives truncation after +8
aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1426c"
/protein_id="CAB73850.1"
/db_xref="GI:6968855"
/translation="MKKEQLEVLKEALNGLIYQIDNHQQAARKPIADVERITSEQLVLN
HWNQENYNALHIOGAIADPOGLGRNLNLDNARIPOKYDIFDKMLVYVYKGGD
GYRVCIDCGGHAGLITDIIILHCGGSQSYIEPNNIYLANFLRKKYENNINVKLFQKAYS
DRNYETDFIMFGNRIILSQGNRIVESQDSQTEKYIKVQVIDLCEFINEILTOHKRIY
FLKLDIGMEFEIMKKIIEKRIKIDYIACETHEYMFDDSEKKIGELKQLINKCNIO
NILDDWI"
138604..138613
/ note="polymorphism C(10-11)"
complement(138926..139867)
/ gene="Cj1427c"
complement(138926..139867)
/ gene="Cj1427c"
note="Cj1427c, probable sugar-nucleotide
epimerase/dehydratase, len: 313 aa; similar to many
predicted epimerases/dehydratases e.g. TR:Q25130
(EMBL:D25537) HALOCYNTHIA RORETZI (SEA SQUIRT) gene with
similarity to UDPglucose4-epimerases and
3beta-hydroxysteroid dehydrogenase/isomerases HREPIB (241
aa), fasta scores; opt: 585 z-score: 691.2 E(): 3.5e-31,
44.2% identity in 208 aa overlap, and TR:O73960
(EMBL:AP000002) Pyrococcus horikoshii hypothetical
UDP-glucose 4-epimerase (318 aa), fasta scores; opt: 304
z-score: 361.7 E(): 7.9e-13, 26.4% identity in 326 aa
overlap. No Hp ortholog"
/codon_start=1
/transl_table=11
/product="putative sugar-nucleotide
epimerase/dehydratase"
/protein_id="CAB73851.1"
/db_xref="GI:6968856"
/translation="MSKKVLITGGAGYIGSVLTPILLEKGYEVCVIDNLMFQDLSLS
CFHNKFTFNGDAMDENLIRQEVAKADIITPLAALVGAPLCRNPKLAKMNIYEAVK
MISDFASPSQIIFYIPNMSGYIGEKDAMCTEESPLRPISEYIDKVAEQYLLDKGN
CVTFRLATVFGISPRMLDILLVNDFTYRVDKAFIVLFEHFRNRYTHVRDVKGFTH
GIENYDKMGQAYNMGLSSANLTKQLAETIKKYPIDPFYHSANIGEDPKRDYLVSN
TKLEATGWKPDNLELDGIELLRAFRKMKVNRANFN"
complement(139891..140931)
/ gene="fcl"
complement(139891..140931)
/ gene="fcl"
note="Cj1428c, fcl, probable fucose synthetase, len: 346
aa; similar to e.g. FCL_ECOLI fucose synthetase (321 aa),
fasta scores; opt: 562 z-score: 690.0 E(): 4.1e-31, 39.0%
identity in 341 aa overlap. 29.6% identity to HP0045.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative fucose synthetase"
/protein_id="CAB73852.1"
/db_xref="GI:6968857"
/translation="MOTNSKIYTAGHKGTAGTALVENLOKRQNNLVLKTRQELDLVN
QQAFAFKPEEKPEYVFLTAVLPGGAANVAQRAFDIYENLMIONNYTHNSFLANVKL
VFGSGYMPENAKNPKLEYLFQGDLEKAYSGFGAAK IAGAIMCSYNIQYGTNFT
LVNLNLYGTKANEDFGKRVLPALLRFLHKLISEGNITQILODLMMNFEAKAYL
HNFGISKKSVEIHWGTGTSDDLAIVTYMQNDIKDILKDRKSKNTHINIG
TGDISYIKEVALMVKNIVGFSGLVFNTPSPDSTMRLMDCSKIHSLGWKHKTELKDG
IKMWEYWKTON"
complement(140398..140421)
/ gene="fcl"
misc_feature

```

ppsa_STAMA probable phosphoenolpyruvate synthase (834 aa),
fasta scores; opt: 166 z-score: 182.8 E(): 0.0074, 23.9%
identity in 226 aa overlap. No Hp ortholog"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1418c"
/protein_id="CAB73842.1"
/db_xref="GI:6968847"
/translation="MAELKFKTKAQNKLQTKLKAKVLPLVLTSLSEELISNEDKVL
QDQITKANRLIIRSSLSDESMKSNAGAFSLANIKADSKDELKALTYEVANSWPS
KSEILLVOPMLENITLCGVGFSDKONFSYFCIQDENGSSNITDGSKSARTYTH
YRNYLEFDRILQKIHILIKLEVLDFDCHDFEFAFAIQDDEELFCQVRPLVME
KNLFLSPREALYRFYKRELSKESRSRVLGDAIFGVMDPWNPAAEIIIGLRPKRLAF
SLYKEIITDWAYQRDNYGRDLRSHPLHSFLGIPYVDVRLSFNSFIKKIDENIA
OKLVNFYLDKLNKHELHDKIEFNIVYSCVDNSKKLELLAHNGENENIKLEPSL
LELTNKIINPRSGFYLDQKAYKALKERYDGIINSFSLIDKIKYLIBECKRGITPLP
AGVARAVAMQNLNLSVEIDFTEKEKDQFLNSLNTVSKNLSQTKNHLNHNKDQFL
KDFGRAGYNTLSIPRYDEFLYFDADQDKSVYLODKAFYSKEFKETRALNALKE
HGLEINACEFDFLQKAREGLVKFETRLLSKAIWYIBELGKYDIEKEDLAHLDI
KSLINLYSSLYSINPKQOFTIEINRNKKEYELTQAIKLPSLLCNADIEIFSYNHSIIP
NFTQKSITAFATAKENDKLEGVILVIYAADPGYDYLFTKNIAGLITCYGGSANHMAI
RASELGMNAVIGVGEENFEKYLKAKKINIECESEQIFCL"
/complement(130296. .131057)
/gene="Cj1419c"
/complement(130296. .131057)
/note="Cj1419c, possible methyltransferase, len: 253 aa;
similar to hypothetical proteins e.g. FR:O53392
(EMBL:AL021841) Mycobacterium tuberculosis Rv3342
(MW016.42) (243 aa), fasta scores; opt: 210 z-score:
250.1 E(): 1.3e-06, 32.0% identity in 122 aa overlap. Also
contains weak similarity in part to several
methyltransferases e.g. UBIE_ECOLI ubiquinone/menaquinone
biosynthesis methyltransferase (251 aa), fasta scores; 132
opt: 141 z-score: 171.3 E(): 0.032, 24.2% identity in 132
aa overlap. No Hp match. Also similar to Cj1420c (33.7%
identity in 255 aa overlap)"
/codon_start=1
/transl_table=11
/product="possible methyltransferase"
/protein_id="CAB73843.1"
/db_xref="GI:6968848"
/translation="MQKLVQVWDYTKHAKFYSRPNYAPKTIIDLISLVGKKDIKVA
DIGAGTGNLSIMLLERGCKVSVPEPDAMREIGERTKQKIDWRATGLNSTLQNSE
FDWVTGSSPNVMDRNEALEARHLKSEGYFSCMNHRDLNDPVOKIAEDTIVEVP
NYTRGTRERQDRPILESKDLFDNIVIEREDFYFHQSNIENAWKSVKNPYWDLTD
EGNLETKSNKLSQRUPKFSITYTRCWSAKKI"
/complement(131110. .131883)
/gene="Cj1420c"
/complement(131110. .131883)
/note="Cj1420c, unknown, len: 258 aa; similar to
hypothetical proteins e.g. TR:O53392 (EMBL:AL021841)
Mycobacterium tuberculosis Rv3342 (MW016.42) (243 aa),
fasta scores; opt: 219 z-score: 263.3 E(): 2.4e-07, 30.7%
identity in 137 aa overlap. Also similar to Cj1419c (34.1%
identity in 255 aa overlap). Contains C(9-10) polymorphic
region at aa 132. C9 (consensus) gives this ORF, C10 would
give truncation after a further 19 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1420c"
/protein_id="CAB73844.1"
/db_xref="GI:6968849"
/translation="MKQGDFTKVAKHYNRPAYSPFLEKLKLVACTINDKNKNLKDNLIV
EVAGCTKLTMLGEMFGQIISAEVNDNNRREGQFTQNLNSIWHKSGSEETCMSN
NOADMIVMASSFWHTDPKSLPEFNRLTGGGGYFTAIWNPRIHVEGSVFDETEKEIKH
IVPDLARVSSGTONVKVWEILVTGDFIDCFPMEDCYKEFWDKERYLGAHWSVNDIQ
AQAGEKRWKEILBMEIAKISHMQSIEIPYKIRAWTARKA"
/complement(131482. .131490)
/note="polymorphism C(9-10)"
131868. .160218
/note="region with low G+C: 26.5%"

complement(131950. .133788)
/gene="Cj1421c"
complement(131950. .133788)
/gene="Cj1421c"
/note="Cj1421c, possible sugar transferase, len: 612 aa;
no Hp match. Similar to polymorphic tract-containing genes
Cj1422c (overall 58.2% identity in 636 aa overlap, but N-
and C-terminal 250 and 40 aa are identical), Cj0685c
(28.0% identity in 404 aa overlap). N-terminus is similar
to parts of Cj1431c (38.6% identity in 114 aa overlap),
and putative sugar transferases Cj1434c (39.6% identity in
111 aa overlap), and Cj1438c (34.6% identity in 127 aa
overlap). Contains C(8-10) polymorphic tract at aa 30. C9
(consensus) gives this ORF. C8 gives truncation after +1
aa. C10 gives truncation after +5 aa. Putative functional
assignment is based on family clustering (BLASTP) with
other C. jejuni predicted sugar transferases e.g. Cj1434c,
and Cj1438c"
/codon_start=1
/transl_table=11
/product="possible sugar transferase"
/protein_id="CAB73845.1"
/db_xref="GI:6968850"
/translation="MLNPNSAIERVKNHLAYKLGQTVIEHRHNGGYIALFKKLYKIK
KQHKKEQIYQOIIQVFPQPKYPSLETCDYNEALRCFKHLSYMGIVLKAQYNNYK
GGGFKLNKNIKRAKQFOIFREILKEFNGETLKAIODNKOLFKEFPRIKNILKT
HODYQPILDNIPIHNFNVIKTNFDLIEWLLSDDFEKYKKNHYPYSLDPPKILNDEN
EXINYHNIPAEALAWKNLPLPPNYEPMWFESHGAGFTLGOFFYHLEKINILDYFCGG
GDRIYIKYFKLLELAKRNIITINDIDPSYGNQHKRDLKFLSSFKQITPILFIQND
PIELIKHAYGRKWNLAKTKEFDLSYQPNDIITEVEYVYNLPNTLEGORPOSFLWK
SLTECFDKNDCFYLDISKIRGEETIHTLNYSKNFLKQIKINDKEFTVTSYKFGNL
FLLPULTLNKLKEDLNTNPKINKNNSLIINFFONDNFLNFYSELSILDMOSS
VGFYIDQYQNLKNDISIFKQVIDYLRNFAYELKRIQIEEDLMKLVEDVLRHLYNN
KNARVSAKNILDIELVYIKQHRPDIIVASWKYQEFPEKMKCKELDGI"
/complement(133693. .133701)
/note="polymorphism C(8-10)"
complement(133843. .135720)
/gene="Cj1422c"
complement(133843. .135720)
/note="Cj1422c, possible sugar transferase, len: 625 aa;
no Hp match. Similar to polymorphic tract-containing genes
Cj1421c (overall 58.2% identity in 636 aa overlap, but N-
and C-terminal 250 and 40 aa are identical), Cj0685c
(25.6% identity in 442 aa overlap). N-terminus is similar
to parts of Cj1431c (38.6% identity in 114 aa overlap),
and putative sugar transferases Cj1434c (39.6% identity in
111 aa overlap), and Cj1438c (23.5% identity in 452 aa
overlap). Contains C(9-10) polymorphic tract at aa 30. C9
(consensus) gives this ORF. C10 gives truncation after +5
aa. Putative functional assignment is based on family
clustering (BLASTP) with other C. jejuni predicted sugar
transferases e.g. Cj1434c, and Cj1438c"
/codon_start=1
/transl_table=11
/product="possible sugar transferase"
/protein_id="CAB73846.1"
/db_xref="GI:6968851"
/translation="MLNPNSAIERVKNHLAYKLGQTVIEHRHNGGYIALFKKLYKIK
KQHKKEQIYQOIIQVFPQPKYPSLETCDYNEALRCFKHLSYMGIVLKAQYNNYK
GGGFKLNKNIKRAKQFOIFREILKEFNGETLKAIODNKOLFKEFPRIKNILKT
HODYQPILDNIPIHNFNVIKTNFDLIEWLLSDDFEKYKKNHYPYSLDPPKILNDEN
EXINYHNIPAEALAWKNLPLPPNYEPMWFESHGAGFTLGOFFYHLEKINILDYFCGG
GKRIYISIFLSEACACNDLKAIFGILDVNCQDKYFCFLQNKQPVLLILDRPIDSL
KSFINYRHMKNFNEILKIDINNTDFKIDIRIVYVHESNGCFNPDTNKKFPPLESIK
ALSDTWHQMLNIRNKTIEFFRNKIIYIDMDIVGDKTLFTLEKLSKLFNSFPD
KNNKIFYOOLYPLTVDLLPCIKVNNKVKIFYSNRSVKNIQIMENICIDITDKFKEIF
HENLILFCSKDFEDSLNNQTLNVLVLEYNKFLISLKKRINVKKVKVDVLDYF
KKNISVAKSYKDLIDELVYIKQHRPDIIVASWKYQEFPEKMKCKELDENQNQNSLSFSN
Q"

misc_difference 135625. .135633
/note="polymorphism C(9-10)"
complement(135771. .136436)
gene

/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73836.1"
/db_xref="GI:6968841"
/translation="MDREVFYIAGYDPKSYFYDLEFKNNLKDYSHAFNIKADLSKIE
KNEQFPFQISCEGVQTKYHFLTWNDIVKKNWSENKDALADCVSPFRIVTITGLFIK
KESIGVQLTGYYPPFYVLFSLFLSVLAFGSPAFQLQNYWHFSLAIIGCGFLIN
HFLKLGKKAIVFARIACAFCACTQDKKTQMOERIKRNFANTIVLKLQKQENKDYE
LVVAHSVGTIVCEVLECIQRONLDSLSLHLKILTLGECIPLVSOKKADFRKKL
EVSFRDLKWDTYSIIDGACFPQVDFRYSVGNAKETPPFLSAKFHTLYEKHEVKKI
KRDKNAHFLYLSISVKGQYDFESFIIMPFLFEKVKI"
complement(122837..124021)
/gene="Cj1413c"
/complement(122837..124021)
/note="Cj1413c"
/protein_id="CAB73837.1"
/db_xref="GI:6968842"
/translation="MRFSTIKIKFEKSGKVLNLLQGVNPFPHLAVKMRKNOTKVFKL
NFNGDFFPSPGTRCKDEKLENFYRDPONKKIDAILMYNDCRIHAKAIKAYKE
LGSIWIFEGLRYPICITLEKQGVNANSSLPDRKNFYLSQNIKESIKRIPGGFKF
MAFADLYWFAFIALPFFNNKLLHRTLYPPEFLFRSLYKRYLKITKLEKRIY
NLKRYFLAILQVSDYQIKYHYKKSIEHTEETILSFANHARAKSYLVFKHHPMDRG
YKNFYKLINDLSRKYHVGEVRLYVHDTHLPVLLRKALGCITINSTVGLSALLECPK
VCGNAFYNEGLSPKQLQFPWEAHAYKPNPVLVCNFKYLLQTNQNGNFYKNFEL
DK"
complement(124018..126087)
/gene="Cj1414c"
/complement(124018..126087)
/note="Cj1414c"
/protein_id="CAB73838.1"
/db_xref="GI:6968843"
/translation="MKYYSISKLIANVRNFTYISLYKKNLKKKDDLFTGWBKKS
LKANNLAKKYKAKFILLEDFIRSLNCGVNSPFSVMDKDDIGIYYDATMPSKLENLL
NTCFKDEELQAKKALIELIKYISKYNNNDIPDPYQFDERKRVLLITQTANDASL
EFGAKDFKLDMDIKDAIKENPKSTIYKICHPDVLGSKQSDLDNLSPKECILITEN
FNPLALFEFDKVTYSGMGFEALQCEKICYGMFPYAGWGLTKDLECKRRMOKR
SLEVFYAAYTLYAEYENPYNOKSNIFDTTOTLAKYDIEKYNKLMFLGFTLWKR
HFIPFKAKDNELIFELNSLKLARKYKLENDKFEFIMGRIDYNALKTLLKKAODEN
LLHPTPKLSLVEDFIRSIGSLDTRPFSLLVDDKGLIDPNKASKLEELLQNEIFD
ENMLNRKNIITLLENRFSYKNGKHELDUKINAKIGOKIILIPAQVEDDASMLGGF
GLSTLDLHTEVRANKQDAYIIFKHPDVLGSGNRVGLKDETILILEFCFVKDCSIDSA
IKIADEHTITSTSGFDALLRAKRVFTYGMFPYAGWGLTKDKYRCERRIRKLSLEELV
AGALIIVPRYINPTKTLCETEVCIDIMNLQKAYFSKKYIKLAIDFKTFMLKIRRF
YEFLAKK"
complement(126084..126596)
/gene="cysC"
/complement(126084..126596)
/gene="cysC"
/EC_number="2.7.1.25"
/note="Cj1415c, cysC, possible adenylylsulfate kinase,
len: 170 aa; similar to e.g. CysC_ECOLI P23846

/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73836.1"
/db_xref="GI:6968841"
/translation="MDREVFYIAGYDPKSYFYDLEFKNNLKDYSHAFNIKADLSKIE
KNEQFPFQISCEGVQTKYHFLTWNDIVKKNWSENKDALADCVSPFRIVTITGLFIK
KESIGVQLTGYYPPFYVLFSLFLSVLAFGSPAFQLQNYWHFSLAIIGCGFLIN
HFLKLGKKAIVFARIACAFCACTQDKKTQMOERIKRNFANTIVLKLQKQENKDYE
LVVAHSVGTIVCEVLECIQRONLDSLSLHLKILTLGECIPLVSOKKADFRKKL
EVSFRDLKWDTYSIIDGACFPQVDFRYSVGNAKETPPFLSAKFHTLYEKHEVKKI
KRDKNAHFLYLSISVKGQYDFESFIIMPFLFEKVKI"
complement(122837..124021)
/gene="Cj1413c"
/complement(122837..124021)
/note="Cj1413c, possible polysaccharide modification
protein, len: 394 aa; similar to e.g. LipB_NEIME caps
polysaccharide modification protein (419 aa), fasta
scores: 897.9 E(): 0, 38.5% identity in
395 aa overlap, and KSS5_ECOLI capsule polysaccharide
export protein KPSS (389 aa), fasta scores: opt: 919
z-score: 1108.7 E(): 0, 40.2% identity in 383 aa overlap.
No Hp match"
/codon_start=1
/transl_table=11
/product="possible polysaccharide modification protein"
/protein_id="CAB73837.1"
/db_xref="GI:6968842"
/translation="MRFSTIKIKFEKSGKVLNLLQGVNPFPHLAVKMRKNOTKVFKL
NFNGDFFPSPGTRCKDEKLENFYRDPONKKIDAILMYNDCRIHAKAIKAYKE
LGSIWIFEGLRYPICITLEKQGVNANSSLPDRKNFYLSQNIKESIKRIPGGFKF
MAFADLYWFAFIALPFFNNKLLHRTLYPPEFLFRSLYKRYLKITKLEKRIY
NLKRYFLAILQVSDYQIKYHYKKSIEHTEETILSFANHARAKSYLVFKHHPMDRG
YKNFYKLINDLSRKYHVGEVRLYVHDTHLPVLLRKALGCITINSTVGLSALLECPK
VCGNAFYNEGLSPKQLQFPWEAHAYKPNPVLVCNFKYLLQTNQNGNFYKNFEL
DK"
complement(124018..126087)
/gene="Cj1414c"
/complement(124018..126087)
/note="Cj1414c, possible polysaccharide modification
protein, len: 689 aa; similar to e.g. LipA_NEIME caps
polysaccharide modification protein (400 aa), fasta
scores: opt: 927 z-score: 1062.5 E(): 0, 39.6% identity in
394 aa overlap, and KSC5_ECOLI capsule polysaccharide
export protein KPSC (675 aa), fasta scores: opt: 724
z-score: 827.0 E(): 0, 38.5% identity in 655 aa overlap.
No Hp match"
/codon_start=1
/transl_table=11
/product="possible polysaccharide modification protein"
/protein_id="CAB73838.1"
/db_xref="GI:6968843"
/translation="MKYYSISKLIANVRNFTYISLYKKNLKKKDDLFTGWBKKS
LKANNLAKKYKAKFILLEDFIRSLNCGVNSPFSVMDKDDIGIYYDATMPSKLENLL
NTCFKDEELQAKKALIELIKYISKYNNNDIPDPYQFDERKRVLLITQTANDASL
EFGAKDFKLDMDIKDAIKENPKSTIYKICHPDVLGSKQSDLDNLSPKECILITEN
FNPLALFEFDKVTYSGMGFEALQCEKICYGMFPYAGWGLTKDLECKRRMOKR
SLEVFYAAYTLYAEYENPYNOKSNIFDTTOTLAKYDIEKYNKLMFLGFTLWKR
HFIPFKAKDNELIFELNSLKLARKYKLENDKFEFIMGRIDYNALKTLLKKAODEN
LLHPTPKLSLVEDFIRSIGSLDTRPFSLLVDDKGLIDPNKASKLEELLQNEIFD
ENMLNRKNIITLLENRFSYKNGKHELDUKINAKIGOKIILIPAQVEDDASMLGGF
GLSTLDLHTEVRANKQDAYIIFKHPDVLGSGNRVGLKDETILILEFCFVKDCSIDSA
IKIADEHTITSTSGFDALLRAKRVFTYGMFPYAGWGLTKDKYRCERRIRKLSLEELV
AGALIIVPRYINPTKTLCETEVCIDIMNLQKAYFSKKYIKLAIDFKTFMLKIRRF
YEFLAKK"
complement(126084..126596)
/gene="cysC"
/complement(126084..126596)
/gene="cysC"
/EC_number="2.7.1.25"
/note="Cj1415c, cysC, possible adenylylsulfate kinase,
len: 170 aa; similar to e.g. CysC_ECOLI P23846

adenylylsulfate kinase (EC 2.7.1.25) (200 aa), blastp
scores: E = 4.2e-13, 28% identity in 164 aa overlap, and
part of NODQ_RHI83 probable sulfate adenylylate transferase
subunit 1 (646 aa), fasta scores; opt: 314 z-score: 373.9
E(): 1.7e-13, 32.9% identity in 161 aa overlap. No Hp
ortholog. Contains PS00017 ATP/GTP-binding site motif A
(P-loop).
/codon_start=1
/transl_table=11
/product="possible adenylylsulfate kinase"
/protein_id="CAB73839.1"
/db_xref="GI:6968844"
/translation="MKNNPYIIWLTGLAGSGKTTIGQALYEKLLKVKYNNLIYLDGDEL
REILGHYAYDRQRIDMALRRFAKELNDQGMVIVTTISMFNEIYDNRKQLKNIY
EYIETCDMHELIOQRDQKGLYTKALNIDNVGVVDIEFDKPEADLVINNSCRNNLEEK
VELTIKKLAL"
complement(126540..126563)
/gene="cysC"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(126586..127347)
/gene="Cj1416c"
complement(126586..127347)
/gene="Cj1416c"
/note="Cj1416c, probable sugar nucleotidyltransferase,
len: 253 aa; similar to predicted nucleotidyltransferases
e.g. TR:O29123 (EMBL:AE001025) Archaeoglobus fulgidus
predicted glucose-1-phosphate cytidylyltransferase AF1142
(241 aa), fasta scores; opt: 252 z-score: 309.2 E():
6.7e-10, 26.7% identity in 251 aa overlap. Also similar in
part to LICC_HAEIN LICC protein (involved in LPS
synthesis) (198 aa), fasta scores: opt: 209 z-score:
259.6 E(): 3.9e-07, 32.7% identity in 107 aa overlap. No
Hp ortholog. Contains Pfam match to entry PF00483
NTP transferase, Nucleotidyl transferase"
/codon_start=1
/transl_table=11
/product="putative sugar nucleotidyltransferase"
/protein_id="CAB73840.1"
/db_xref="GI:6968845"
/translation="MNAIILAAAGFGSRLMPLTKDQKCMVEYKNNKIIDYEIEALKSA
GINIAVVGVLNDVLNKLNYDIEHFITNSKYDKTMVHTPFCAKDFMLCKLEEKQ
DLIISYADIVYFQDCVOKLINAKELAIIVDKSKLWKRFPANPLEDAETLKWNTGY
ITELGKRANAYDEEAQYIGLFKFSYQFLSEVYAFYEMLDRLDILYDNKNFNMYMTSF
LQALIERYNNAKAVEIDGNCEIDFMSDLEVQIEK"
complement(126598..127344)
/gene="Cj1416c"
/note="Pfam match to entry PF00483 NTP transferase,
Nucleotidyl transferase, score -74.20, E-value 0.016"
complement(127349..127951)
/gene="Cj1417c"
complement(127349..127951)
/gene="Cj1417c"
/note="Cj1417c, unknown, len: 200 aa; some similarity to
the C-terminus of carbamoyl-phosphate synthase small
subunits e.g. CARX_BACSU carbamoyl-phosphate synthase,
arginine-specific, small chain (353 aa), fasta scores;
opt: 134 z-score: 174.1 E(): 0.022, 27.3% identity in 99
aa overlap. No Hp ortholog"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1417c"
/protein_id="CAB73841.1"
/db_xref="GI:6968846"
/translation="MFTIGTQRLICNDYSYHEKRECLADMGKLFNKDIFKNFTPLPLS
VEIDFSYVHLIKAVILSGNDLSFSPNVLKRDLYEKQVIECLEEKIPLGICR
GAOMIAHYENSHISPCENHICKHEVFESKFKFISNPFNAIEKLGEDLVCLAKDN
TTEAFKHYNIFIGIMWHIERENLNIIQLKEWFSLIKE"
complement(127942..130281)
/gene="Cj1418c"
complement(127942..130281)
/gene="Cj1418c"
/note="Cj1418c, unknown, len: 779 aa; weak similarity in
part to several phosphoenolpyruvate synthases e.g.

similar to YQJC_ECOLI (predicted periplasmic protein) (127 aa), fasta scores; opt: 160 z-score: 185.2 E(): 0.0054
32.4% identity in 111 aa overlap. No Hp match. Contains probable N-terminal signal sequence"

```

/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73830.1"
/db_xref="GI:6968835"
/translation="MKKILILTLCAFAFGACSECDRIKIDINKEIFSKAHNDTARTL
SLEIALKQVNDCAKPMFYDKKLEAKKLEQEVEKIEKDALKEQKDYMSAEYKA
KKEALKQEKIRK"
complement(117494. .118882)
/gene="Cj1407c"
CDS
complement(117494. .118882)
/gene="Cj1407c"
/note="Cj1407c, probable phospho-sugar mutase, len: 462
aa; simial to e.g. ALGC_PSEAE phosphomannomutase (462 aa),
fasta scores; opt: 1037 z-score: 1216.2 E(): 0. 40.7%
identity in 457 aa overlap, and PGMU_NEIGO
phosphoglucomutase (460 aa), fasta scores; opt: 884
z-score: 1037.2 E(): 0. 35.4% identity in 435 aa overlap.
44.7% identity to HPI275. Contains Pfam match to entry
PF00408 PGM_PMM, Phosphoglucomutase and
phosphomannomutase, and PS00017 ATP/GTP-binding site motif
A (P-loop). Also similar to Cj0360 (28.8% identity in 386
aa overlap)"
/codon_start=1
/transl_table=11
/product="putative phospho-sugar mutase"
/protein_id="CAB73831.1"
/db_xref="GI:6968836"
/translation="MNLKMKLDVIFREYDIRGLYDKELNEKSKVATGFCIGOTMLNR
GCRNVSGYDARYSANELPNLYSGLNKAGIKYDGLVPTPLGYFSLYSGLFADNV
MITGSHNPDIYNGFKITINKESFSGVELKEFSKEVIKHLDDIEENLEVEKIDILNL
VKMEQESFLKDFNYKFGVDYNGAAGVVIKPLIKALNLKAHVMEAPDQFPNHP
DPTTEENLSAIRFLNQNDYSLAFADGDADRMVALSKTHVFCGDELCLYLFADIPN
PRILGEVSKNLFDEVAKEGFTFMGKTGHSNTKMMKEDIIDLAAEVSGHIFFKRY
FGYDGIYALRALLEVYKGFDSLEMTKALPKLYTPEIKIPVNEEEKFLVBEFOKE
IEKGALGVSLCEIDGARIDFGDWALLRASNTSPVLIIRFATSLSERAKELSMVF
TLFNDIKARKFN"
complement(117554. .118861)
/misc_feature
/note="Pfam match to entry PF00408 PGM_PMM,
Phosphoglucomutase and phosphomannomutase, score 295.00,
E-value 9.3e-85"
complement(117971. .117994)
/misc_feature
/note="Cj1407c"
/translation="MDELENEETKKKGGSLVIIIVILLFVLLSLTMGVIAWLISS
SSDEVEKAPKEAKADPKVSAQAQSDGFANIGMPLDPTFLNLLSDSGSRVYK
CTLEQNSLELLPELDKKVPVIRDIIRTLTAKTEEVSTQKGERLKDELGVKINE
ILTDGFKNYFTDFVVS"
119521. .119868
/gene="acps"
CDS
119521. .119868
/EC_number="2.7.8.7"

```

```

/note="Cj1409, acps, probable holo-[acyl-carrier protein]
synthase, len: 115 aa; similar to e.g. ACPS_ECOLI
holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (125
aa), fasta scores; opt: 170 z-score: 237.9 E(): 6.3e-06,
29.4% identity in 119 aa overlap. 51.3% identity to
HP0808"
/codon_start=1
/transl_table=11
/product="putative holo-[acyl-carrier protein] synthase"
/protein_id="CAB73833.1"
/db_xref="GI:6968838"
/translation="MRVGCDDIIAISRIEKIHSRHGKNFLDKFLNPKQIILKNPATIA
GLWAAKEAASKALGVICELCSFDEISKDEKNAPLKYSQKITKDFENITQTSLSIS
HDNGFAIAIVAVV"
complement(119869. .120288)
/gene="Cj1410c"
CDS
complement(119869. .120288)
/gene="Cj1410c"
/note="Cj1410c, possible membrane protein, len: 139 aa; no
Hp match. Contains hydrophobic domain and PS00041
Bacterial regulatory proteins, araC family signature"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73834.1"
/db_xref="GI:6968839"
/translation="MKRVRTKIRANFRVRKRTLKSLKEKLACTILLCAIVPLAVLG
YLFIVIGFIFFNARQGVRAIDHFNVASLFGYAWESVSSHWARNRKKWARIYI
KITDFQDKHCKRANKREQVPVDFILSRNLDKQTIGK"
complement(120112. .120246)
/gene="Cj1410c"
/EC_number="PS00041 Bacterial regulatory proteins, araC family
signature"
complement(120285. .121646)
/gene="Cj1411c"
CDS
complement(120285. .121646)
/gene="Cj1411c"
/note="Cj1411c, probable cytochrome p450, len: 453 aa;
similar to many e.g. cytochrome p450 domain of CPXB_BACME
cytochrome P450/NADPH-cytochrome P450 reductase (1048 aa),
fasta scores; opt: 505 z-score: 578.8 E(): 6.4e-25, 28.3%
identity in 407 aa overlap. No Hp match. Contains PS00086
Cytochrome p450 cysteine heme-iron ligand signature, and
Pfam match to entry PF00067 p450, Cytochrome p450"
/codon_start=1
/transl_table=11
/product="putative cytochrome P450"
/protein_id="CAB73835.1"
/db_xref="GI:6968840"
/translation="MSBCPPFPKPKYKNAKSTLLTFLTKRRSLDGLGYERSYKMQTGVY
KMPNFDLYINDTKVKRMVDEVREPKSAFLHELLSPILGSIPTTNGEVMKKORE
LLRPSFEMTRINKVFNLMSEAVADMDDRFKSKYPNHAVIDEAMTFITADVIERTIMS
SKLDEEGRKKILNAFVTFQGSVHTAMRRMFRKFSYVILGDCRAKAGDVIRQVUS
DIIKRPMADNAEFEDILGSLLLVADNRKRFSEFIEDQVAMFLAGHETASSL
TWLYLLSLYPKEQKAYEITVLAQGVIEIHLRQFVLTNIFKEISLYPPGVFF
ARAKKQTOVRDKLIRKKGSGVIAPLIHRHEEFTWPHGFNPSRPFGEYKKDAYLPF
GVGERICIGGFAMQEAILILANILKTYKLEBEGFVPDVVGRITVRSANGMRKFKSK
RKL"
complement(120303. .121592)
/misc_feature
/genes="Cj1411c"
/note="Pfam match to entry PF00067 p450, Cytochrome p450,
score 216.60, E-value 3.7e-61"
complement(120444. .120473)
/misc_feature
/genes="Cj1411c"
/note="PS00086 Cytochrome p450 cysteine heme-iron ligand
signature"
complement(121643. .122764)
/gene="Cj1412c"
CDS
complement(121643. .122764)
/gene="Cj1412c"
/note="Cj1412c, probable integral membrane protein, len:
373 aa; no Hp match"
/codon_start=1

```

```

/misc_feature
/translation="MIMKGKKLIVGVANNKSIAYGIAKACFDGCAELAFITLNDALK
KRVEPIAOEENSFVYELDANNNEHLDSTAEKIKKDLGEIDEVYHAFAPAKEALENS
ELETSEKAFDIAMOTSVYLSLSLRAVLPIILKDKGSILUTLSYLGVGKVKYPHYNVMGVA
KAALESSYMLARDLUGVKGIRVNAISAGPIKTLAASGIGDFRMILKYNEINSPKRNRY
SIEDVGSAMYLSDSLANGVTGEIHYVDAGYNIMGDGVKNEEQGTVLCWDNQKG"
complement(112471. .112563)
/gene="fabI"
/note="pfam match to entry PF00678 adh_short_C2, Short
chain dehydrogenase/reductase C-terminus, score 35.80,
E-value 9.9e-07"
complement(112627. .113190)
/gene="fabI"
/note="pfam match to entry PF00106 adh_short.
Alcohol/Other dehydrogenases, short chain type; score
49.50, E-value 7.3e-11"
complement(113209. .113880)
/gene="tpiA"
complement(113209. .113880)
/gene="tpiA"
/EC_number="5.3.1.1"
/note="Cj1401c, tpiA, probable triosephosphate isomerase,
len: 223 aa; similar to many e.g. TPIS_MYCHR
triosephosphate isomerase (BC 5.3.1.1) (TIM (243 aa),
fasta scores; Opt: 377 z-score: 466.3 E(): 1.2e-18, 34.5%
identity in 235 aa overlap, 48.2% identity to HP0194.
Contains Pfam match to entry PF00121 TIM, Triosephosphate
isomerase"
/codon_start=1
/transl_table=11
/product="putative triosephosphate isomerase"
/protein_id="CAB73825.1"
/db_xref="GI:6968830"
/translation="MIFAANLKCNHTRASFKIYAKILNKTGMVKCDIIVPPPSIAFL
ENENFIQGAQNFYCVNGAFBELGKHELDGFKICVLIGHSSRRALGDEEFLKAF
DFAEHGYKIVFCIGENLTKNSGKTLEFLKQLEIIDLNEYKLIAYEPIYISGTGV
SAQSTDIAKVLFLASLTVPVLYGGVNNENIKEILSVNHCYGLIGSAALKYENFI
KLIKG"
complement(113395. .113724)
/gene="tpiA"
/note="pfam match to entry PF00121 TIM, Triosephosphate
isomerase, score 95.60, E-value 3.6e-30"
complement(113877. .115079)
/gene="pgk"
complement(113877. .115079)
/gene="pgk"
/EC_number="2.7.2.3"
/note="Cj1402c, pgk, phosphoglycerate kinase, len: 400 aa;
similar to e.g. PGK_AQUAE phosphoglycerate kinase (EC
2.7.2.3) (397 aa), fasta scores; Opt: 1417 z-score: 1610.9
E(): 0.53.4% identity in 393 aa overlap, and PGKY_WHEAT
phosphoglycerate kinase, cytosolic (401 aa), fasta scores;
Opt: 1369 z-score: 1556.5 E(): 0.53.1% identity in 392 aa
overlap, 49.1% identity to HP1345. Contains PS00111
Phosphoglycerate kinase signature, and Pfam match to entry
PF00162 PGK, Phosphoglycerate kinases"
/codon_start=1
/transl_table=11
/product="Phosphoglycerate kinase"
/protein_id="CAB73826.1"
/db_xref="GI:6968831"
/translation="MSDIIISKIDIDLAKKVFTRCDFNVPPQDDFLNITDDBRRISAIP
TIRYCLONGSVILASHLGRPKELSSKYSLEPVAKRLARLLDREIVNAKDVIGEDAKT
KAMNKAAGEILLLENLRFKEGTNDLAKELASVQVYVINDAFGVCHRAHSVEAI
TKFDEKKGAGFLLOKEIDFASNLIKHPARFPVAVYGVSKVSKLQALNLLPKVDDK
LITGGMAFTFLKALGYDIGNLSLEELBEANKILTKGNLGVKIYLPVDVVAAPAC
SQDPMKFVPAQEIPIPCWGMGLDIGPASVRLFKVEISDAOTIWNNGPMGVPEIDKFSKG
SIKMSHVISEGHATSVVGGGDTADVVARAGADAMETISTGGGASLELIEGKELPGVK
ALRSKENE"
complement(113886. .115079)
/gene="pgk"
/note="pfam match to entry PF00162 PGK, Phosphoglycerate
kinases, score 697.10, E-value 8.2e-206"
complement(115002. .115034)

/misc_feature
/translation="MQRIDILVILDEKKAENIKTTDMSEQEVFVKYVYIAATLGR
LAULSIDLRLKAKGEEFLNDSSEWSVIDGDILIHLLTPHEHRIYNIIELEEN
LKKGRV"
complement(117097. .117447)
/gene="Cj1406c"
complement(117097. .117447)
/gene="Cj1406c"
/note="Cj1406c, probable periplasmic protein, len: 116 aa;

```

```

/gene="pgk"
/note="PS00111 Phosphoglycerate kinase signature"
complement(115082. .116080)
/gene="gapA"
complement(115082. .116080)
/gene="gapA"
/note="Cj1403c, gapA, glyceraldehyde 3-phosphate
dehydrogenase, len: 332 aa; similar to e.g. G3P1_BACSU
glyceraldehyde 3-phosphate dehydrogenase 1 (334 aa), fasta
scores; Opt: 1070 z-score: 1216.4 E(): 0.50.3% identity
in 332 aa overlap, 45.6% identity to HP1346 and 46.7%
identity to HP0921. Contains Pfam match to entry PF00044
gpdh, glyceraldehyde 3-phosphate dehydrogenases, and
PS00071 Glyceraldehyde 3-phosphate dehydrogenase active
site"
/codon_start=1
/transl_table=11
/product="glyceraldehyde 3-phosphate dehydrogenase"
/protein_id="CAB73827.1"
/db_xref="GI:6968832"
/translation="MAVKVAINGFGRIQRCVARILERNIDIELVAINDTTDIETLTKYL
FKYDTVHGFEFGKGSVDGDLVNGKKIKVFKSRNVKDLDFAKGAQVLCTGAHLT
MAKQOEFTDMGVOKVIMSAKADDTPTVVLGVNSELYKGESTIISNASTTNCILCPVCR
VLQDNFGLEKGLMTTIHATYNGQSLIDAKADKRRSRAAQNIIPTSTCAAKAKMLVM
PELNGKLHGKMRVPVIDVSSVDLTAQLSRKVSKEINEAFKAAATNLKGLIMVDDD
ERVSSDFITCSYGAIVASDLTQVIAADDFIKVIAYWIDNEWGYSRLVDMVYIANKA"
complement(115085. .116080)
/gene="gapA"
/note="pfam match to entry PF00044 gpdh, glyceraldehyde
3-phosphate dehydrogenases, score 576.10, E-value
2.2e-169"
complement(115616. .115639)
/gene="gapA"
/note="PS00071 Glyceraldehyde 3-phosphate dehydrogenase
active site"
116139. .116684
/gene="Cj1404"
116139. .116684
/gene="Cj1404"
/note="Cj1404, unknown, len: 181 aa; similar to
hypothetical proteins e.g. YOEJ_BACSU (189 aa), fasta
scores; Opt: 305 z-score: 391.3 E(): 1.8e-14, 30.2%
identity in 182 aa overlap, 40.4% identity to HP1337"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1404"
/protein_id="CAB73828.1"
/db_xref="GI:6968833"
/translation="MKIALFGSDPPHNGHNSVVALEKLDIKLIIMPTYINPFK
QSFSADEKQRFWMYKWLHGLHPKVEICDFEIRQRPVPSPYESVLYKLYNPSKPYLL
IGADHLEKLHLHDFEKLNSLVEFVIANRNDIGIPKNFKDLTKNKKIASSFIRDTLNT
NEVCEEIKDEVKYVEKLOKN"
116712. .117038
/gene="Cj1405"
116712. .117038
/gene="Cj1405"
/note="Cj1405, unknown, len: 108 aa; similar to
hypothetical proteins e.g. YBBB_ECOLI (105 aa), fasta
scores; Opt: 185 z-score: 245.5 E(): 2.4e-06, 28.1%
identity in 96 aa overlap, 51.4% identity to HP1414"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1405"
/protein_id="CAB73829.1"
/db_xref="GI:6968834"
/translation="MQERIDILVILDEKKAENIKTTDMSEQEVFVKYVYIAATLGR
LAULSIDLRLKAKGEEFLNDSSEWSVIDGDILIHLLTPHEHRIYNIIELEEN
LKKGRV"
complement(117097. .117447)
/gene="Cj1406c"
complement(117097. .117447)
/gene="Cj1406c"
/note="Cj1406c, probable periplasmic protein, len: 116 aa;

```

gene

106027..107394
/gene="Cj1394"
CDS

/note="Cj1394"
similar to e.g. PUR8_BACSU adenylosuccinate lyase (431 aa), fasta scores; opt: 727 z-score: 829.8 E(): 0, 33.3% identity in 423 aa overlap, and PCAB_PSEPU 3-carboxy-cis,cis-muconate cycloisomerase (407 aa), fasta scores; opt: 721 z-score: 823.4 E(): 0, 32.8% identity in 399 aa overlap. No Hp ortholog. Also similar to Cj0023 purB (30.2% identity in 440 aa overlap). Contains PS00163 fumarate lyases signature, and Pfam match to entry PF00206 lyase_1, Lyases"
/codon_start=1
/transl_table=11
/product="putative fumarate lyase"
/protein_id="CAB73819.1"
/db_xref="GI:6968825"
/translation="MTGISVFDHRLIADSWSTQEMRAIFCEQNRIOKWLVDLAAAKA QAKLIIPKKADEIAKKAHYKEMDMDFIFAEPKTKHPLVPTVRGLEKACDNNLGEY VHERVTTQDIDLVLOPKKEAMTLVKSELKAIKAKALAKTHKTYTAMGRILAQO LPITFGHKVAIWLSELDHRFEILELEKRLYGVISVGAGTKASLSDECNEKLTLE NLGEPVINSQSTRDRFTIELGVGNINATFNKIAHQHLLILSHNEIDEVAEPFGQO VGSSTPHKRNPAVSENAVTVSNFRANLAILSDIERHEHERDQGVYKMEWKLPPMF LMLSVLANNKFALEVKKDKLNLDTLKGFVLAERVMFALSDDHYGKQHAHEIVY ENAMLGTEQOKTFKEVLLADTRYVSQVLKEKDIDALLDATSYGVYAPKLVDLEFLAKIN SAILK"
misc_feature
106318..106905
/gene="Cj1394"
/note="Pfam match to entry PF00206 lyase_1, Lyases, score 99.60, E-value 2.3e-27"
106858..106887
/gene="Cj1394"
/note="PS00163 Fumarate lyases signature"
107403..108743
/gene="Cj1395"
107403..108743
/gene="Cj1395"
/note="Cj1395, possible pseudogene, len: 1338 bp; similar to hypothetical proteins e.g. YXEY_BACSU (445 aa) (approx 25% identity), and IRG1_MOUSE immune-responsive protein 1 (fragment) (646 aa) (approx 30% identity). No Hp match."
/codon_start=1
/pseudo
/transl_table=11
/product="pseudogene (hypothetical protein Cj1395)"
108825..109049
/gene="Cj1397"
108825..109049
/gene="Cj1397"
/note="Cj1397, unknown, len: 74 aa; similar to hypothetical proteins e.g. TR:AD06110 (EMBL:AE001486) H. pylori J99 JHP0533 (78 aa), fasta scores; opt: 143 z-score: 219.1 E(): 6.9e-05, 41.2% identity in 68 aa overlap, and TR:P73181 (EMBL:D90904) Synechocystis sp. srr2333 (79 aa), fasta scores; opt: 94 z-score: 151.5 E(): 0.41, 28.6% identity in 77 aa overlap. Not predicted in TIGR Hp, but 38.8% identity to c(615977)..616192"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1397"
/protein_id="CAB73821.1"
/db_xref="GI:6968826"
/translation="MTLNELKDGQKAIIVNLNAHKELKNRLLSFGFIKNNKKIHHSS LKNATIMVELDTSVILRSDEAKTIEVNLI"
109046..110887
/gene="feob"
109046..110887
/gene="feob"
/note="Cj1398, feob, probable ferrous iron transport protein, len: 613 aa; similar to many e.g. FEOB_ECOLI ferrous iron transport protein B (773 aa), fasta scores;

misc_feature

opt: 726 z-score: 805.2 E(): 0, 29.3% identity in 726 aa overlap. 50.5% identity to HP0687. Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="ferrous iron transport protein"
/protein_id="CAB73822.1"
/db_xref="GI:6968827"
/translation="MKKIKIALVGPVNGKSLILINALCKAMKMGVNFSGVTIEKASAK TFYKNYFEVIDLPGYSLDGYSEEEKITRHFNLQNDYDVIVNVLDTATNLERNLISA ELLSLNKKMLLALNMCDEAKKEGIELDTSILSOFQSVVEISAKYENLELLLOKII ILFESKPIRSPQYTPLCESKPEKDLLYFINELSKLIITHKKEENLTKKIDALLIH KFGLEIFLFWMLLQFLTSQIPMDYTESGFNTLGEFVNKNISNFTTASALADI IAGVGILFPLNIMILFGLIALETLTGYSRVAFLDGLHKGFLHGKGFPLITGFI GCSVPAPMATRTLNKRDRLTLFVINFMSCGARLPVYVLFICAFPPSERAGNYLFGI YILGAILGLCAAKFLBMTAFRGLDEFPVEMPKYRMPNWHLVWFMYNKKAKMYLKKAG TFILLASLLIWFASNFPSKEENLDFNAQERATEQSLYGOFGKIEPFPQLELDWKL SVSLISGLAAKEVIMSTGMVLYSLGKDVDTNNDLKGIIAKNIPISAVAFILFVMYI NPCFAATIVFSKESGKLKVTFLFLEFCTTSAYIVAFIHLIAKILLN"
109073..109096
/gene="feoB"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(110884..112377)
/gene="hyda2"
complement(110884..112377)
/gene="hyda2"
/note="Cj1399c, hyda2, probable Ni/Fe-hydrogenase small subunit, len: 497 aa; similar to e.g. MBHS_WOISU quinone-reactive Ni/Fe-hydrogenase small chain (354 aa), fasta scores; opt: 1122 z-score: 1317.8 E(): 0, 52.6% identity in 285 aa overlap. No Hp ortholog. Also similar to Cj1267c (51.5% identity in 330 aa overlap). Contains C-terminal extension not present in other hydrogenase small subunits"
/codon_start=1
/transl_table=11
/product="putative Ni/Fe-hydrogenase small subunit"
/protein_id="CAB73823.1"
/db_xref="GI:6968828"
/translation="MAKLSNEELKNLEDRICKLENSTLKEDAVINEESVKILARHLS LGNEIPALQRFQPIAPKTKLWMLHCECTGCSELSRLSELPSFDELIFDFFSLEYHE TMAANGTAAEELLEYLEEDFLAVEGGVAADITFTLIGAGESGYELEKLAKA KALFAYGTGSSYGGIQAAYPNPCTGCISEVLQKVVNIPGPPSDINIATLSFAL FGVLPELDQONRPWAYKCLHDMCERKAFESGIFAEHFDDCAKNGACLFKIGCG PYTYNCPKVKNAKTSWPVAAGHGCIACSEKNFWEFGNYEKPMANIFYAKLCNEE LKQEPFLEQIKTLEIQIDFEFFESNIKLIILQNIKALGALLVENVYKSPKYNATIEQ KPKLDVSKSRMTLIYLCGLDFGVAYSTLKAFEDNIARIKISLLKAS"
complement(112381..113205)
/gene="fabI"
complement(112381..113205)
/gene="fabI"
/EC_number="1.3.1.9"
/note="Cj1400c, fabI, probable enoyl-[acyl-carrier-protein] reductase [NADH], len: 274 aa; highly similar to many e.g. FABI_ECOLI enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9) (261 aa), fasta scores; opt: 789 z-score: 906.5 E(): 0, 47.1% identity in 257 aa overlap. 67.9% identity to HP0195. Also similar to Cj0807 (29.2% identity in 257 aa overlap), Cj0485 (28.3% identity in 254 aa overlap), Cj1332 (25.4% identity in 244 aa overlap), and Cj0435 (25.0% identity in 256 aa overlap). Contains pfam match to entry PF00106 adh_short, Alcohol/other dehydrogenases, short chain type, and Pfam match to entry PF00678 adh_short_C2, Short chain dehydrogenase/reductase C-terminus"
/codon_start=1
/transl_table=11
/product="putative enoyl-[acyl-carrier-protein] reductase [NADH]"
/protein_id="CAB73824.1"
/db_xref="GI:6968829"

| | | | |
|--------------|--|--------------|---|
| gene | 101742..102212 /gene="Cj1386" 101742..102212 /gene="Cj1386" /note="Cj1386, ankyrin-repeat containing protein, len: 412 aa; similar to proteins downstream of catalase in Pseudomonas e.g. TR:P95546 (EMBL:U16026) Pseudomonas syringae ANKF precursor (183 aa), fasta scores; opt: 217 z-score: 286.8 E(): 1.2e-08, 31.3% identity in 134 aa overlap, and to eukaryotic ankyrin proteins e.g. ANKI_MOUSE ankyrin (1862 aa), fasta scores; opt: 181 z-score: 226.3 E(): 2.8e-05, 29.1% identity in 117 aa overlap. No Hp match. Also similar to Cj0834c (29.3% identity in 99 aa overlap). Contains 2x Pfam match to entry PF00023 ank, Ank repeat" /codon_start=1 /transl_table=11 /product="ankyrin-repeat containing protein" /protein_id="CAB73813.1" /db_xref="GI:6968820" | misc_feature | 103037..103393 /gene="Cj1388" /note="Pfam match to entry PF01042 DUF10, Domain of unknown function" 103403..104842 /gene="Cj1389" 103403..104842 /gene="Cj1389" /note="Cj1389, probable transmembrane transport protein pseudogene, len: 1437 bp; similar to e.g. YHCL_ECOLI hypothetical transport protein (455 aa) (approx 45% identity), and DCUC_ECOLI c4-dicarboxylate anaerobic carrier (461 aa) (approx 31% identity). No Hp match" /codon_start=1 /pseudo |
| CDS | | gene | 104855..106017 /product="pseudogene (transmembrane transport protein)" 104855..106017 /gene="metC" 104855..104992 /gene="metC" /EC_number="4.4.1.8" /note="Cj1392, metC', probable cystathionine beta-lyase, N-terminus, len: 348 aa; similar to many e.g. METC_HAEIN cystathionine beta-lyase (EC 4.4.1.8) (396 aa), fasta scores; opt: 88 z-score: 151.7 E(): 0.4, 45.2% identity in 42 aa overlap. No Hp match. Note that the rest of metC is in the downstream ORF Cj1393. This may represent a pseudogene fragment, or possibly programmed ribosomal frameshifting" /codon_start=1 /transl_table=11 /product="putative cystathionine beta-lyase, N-terminus" /protein_id="CAB73817.1" /db_xref="GI:6968823" |
| misc_feature | THKGDLSLLAAYNNSESARKLLEKGAQVDEKNDRGOTPLAGVFGYDPMCELLVK YGANIDENNGUGMPTPYTFAINMFGKRQVAEFLKKSNFLKISLKILKFIKKF" 101880..101978 /gene="Cj1386" /note="Pfam match to entry PF00023 ank, Ank repeat, score 31.00, E-value 2.8e-05" 101979..102077 /gene="Cj1386" /note="Pfam match to entry PF00023 ank, Ank repeat, score 23.80, E-value 0.004" complement(102213..102869) /gene="Cj1387c" complement(102213..102869) /note="Cj1387c, unknown, len: 218 aa; similar to hypothetical proteins e.g. YHEQ_ECOLI (240 aa), fasta scores; opt: 164 z-score: 202.0 E(): 0.00063, 29.4% identity in 204 aa overlap. No Hp match. Contains probable helix-turn-helix motif at aa 194-215 (Score 1100, +2.93 SD)" /codon_start=1 /transl_table=11 /product="helix-turn-helix containing protein" /protein_id="CAB73814.1" /db_xref="GI:6968821" | CDS | 104971..106017 /gene="metC" /EC_number="4.4.1.8" /note="Cj1393, metC', probable cystathionine beta-lyase, len: 348 aa; similar to many e.g. METC_ECOLI cystathionine beta-lyase (EC 4.4.1.8) (395 aa), fasta scores; opt: 802 z-score: 979.2 E(): 0, 35.8% identity in 354 aa overlap. 36.2% identity to HP0106. Also similar to Cj1727c, 30.0% identity in 363 aa overlap. Contains PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate attachment site, and Pfam match to entry PF01053 Cys_Met_Meta_PP, Gamma-family of PLP-dependent enzymes. Note that the N-terminal 50 aa of the protein appear to be in the upstream ORF Cj1392. This may represent a pseudogene fragment, or possibly programmed ribosomal frameshifting" /codon_start=1 /transl_table=11 /product="putative cystathionine beta-lyase" /protein_id="CAB73818.1" /db_xref="GI:6968824" |
| gene | 103037..103399 /gene="Cj1388" /note="Cj1388, unknown, len: 120 aa; similar to prokaryotic hypothetical proteins e.g. YA28_PYRHO Pyrococcus horikoshii PHAL028 (137 aa), fasta scores; opt: 399 z-score: 503.4 E(): 1e-20, 53.8% identity in 117 aa overlap, and to e.g. UK14_HUMAN 14.5 kD translational inhibitor protein (137 aa), fasta scores; opt: 384 z-score: 485.1 E(): 1.1e-19, 48.3% identity in 116 aa overlap. 53.0% identity to HP0944. Also similar to Cj0327 (34.4% identity in 96 aa overlap). Contains Pfam match to entry PF01042 DUF10, Domain of unknown function" /codon_start=1 /transl_table=11 /product="hypothetical protein Cj1388" /protein_id="CAB73815.1" /db_xref="GI:6968822" | misc_feature | 104980..106002 /gene="metC" /note="Pfam match to entry PF01053 Cys_Met_Meta_PP, Gamma-family of PLP-dependent enzymes, score 326.30, E-value 3.5e-94" 105445..105489 /gene="metC" /note="PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate attachment site" |
| CDS | HISGRSDSLTAPASFLMKNKKYDFLQDYKALVQKSKLIRGSFFFIKNDKLVG ILCINHTSIMRDLICKMDIKGMDGDLGNISFSQNDSTETLSHSDILQVSV DSSYLNSDYQLSITQKEIAEKLYEGFIENKGAVPPIAKFLKISEPSVRYLKKFKK " 103037..103399 /gene="Cj1388" /note="Cj1388, unknown, len: 120 aa; similar to prokaryotic hypothetical proteins e.g. YA28_PYRHO Pyrococcus horikoshii PHAL028 (137 aa), fasta scores; opt: 399 z-score: 503.4 E(): 1e-20, 53.8% identity in 117 aa overlap, and to e.g. UK14_HUMAN 14.5 kD translational inhibitor protein (137 aa), fasta scores; opt: 384 z-score: 485.1 E(): 1.1e-19, 48.3% identity in 116 aa overlap. 53.0% identity to HP0944. Also similar to Cj0327 (34.4% identity in 96 aa overlap). Contains Pfam match to entry PF01042 DUF10, Domain of unknown function" /codon_start=1 /transl_table=11 /product="hypothetical protein Cj1388" /protein_id="CAB73815.1" /db_xref="GI:6968822" | misc_feature | 104980..106002 /gene="metC" /note="Pfam match to entry PF01053 Cys_Met_Meta_PP, Gamma-family of PLP-dependent enzymes, score 326.30, E-value 3.5e-94" 105445..105489 /gene="metC" /note="PS00868 Cys/Met metabolism enzymes pyridoxal-phosphate attachment site" |

KVLNPLUSEPLKXQKNKFLMFLKNDKFAAFSFLKDAHKYGFGLLSYQRFKLSHQKA
LKLAKELNVFVDEKRLNLYVHLQSLKEIEKNFKFLEKNPYAMLSAHSALARPITWASE
NFCLELGMKMSLNDLFQNGIYFKKGIDFKLOEKNNNQMYEILKKOGIKPEAPYNLYD
FLELDPKSGDNILKLTOKGLVLRHSHNLFIEKQALEKLMQELNLLKQNSLDVQSMK
EYFNLSRKAYAIAYLXLDKFPQVSKREAEKRFITNI"
95442. .96428
/gene="selB"
/note="Pfam match to entry PF00009 GTP_EFTU, Elongation
factor Tu family (contains ATP/GTP binding P-loop), score
69.70, E-value 6.3e-17"
95442. .97247
/gene="selB"
95469. .95492
/gene="selB"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
97296. .98006
/gene="Cj1380"
97296. .98006
/gene="Cj1380"
/note="Cj1380, probable periplasmic protein, len: 236 aa;
some similarity to e.g. DSBG_ECOLI thiol:disulfide
interchange protein DSBG precursor (248 aa), fasta scores;
opt: 138 z-score: 165.8 E(): 0.062, 26.8% identity in 164
aa overlap (the active site is conserved). 34.1% identity
to HP0231. Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73807.1"
/db_xref="GI:6968814"
/translation="MKLSLILVCASLPAASNSEISDFYSKSIKAQFPNATVSVSNR
QKVGNTGPSVIVSVLNGKQENILFTKDSLITPDLIDLK*GISVAQYEMKKFQEA
RENFTKNKAVAQKETMIALGDKNKPAIYVSDPCYCREHLAQIDDELKNYQVNY
ILTPVHGSAPEKSALIYKEARKANDREKIALINKYDANIKNPYKVSDAELKEVFS
LYEKVRSGLSATPTIIK"
98012. .98542
/gene="Cj1381"
98012. .98542
/gene="Cj1381"
/note="Cj1381, probable lipoprotein, len: 176 aa; no Hp
match. Contains probable N-terminal signal sequence and
appropriately positioned PS00013 Prokaryotic membrane
lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAB73808.1"
/db_xref="GI:6968815"
/translation="MSKFAILISFFIALPMSACTNATFTQSFIO*TSNEGIFIRSQKQ
SPKISFONPSOLOTTLDRDLALKNLGLKEVKENADYEILINLVDMKKHSAQKITT
SARFYDFDPLESDGEWVENVYTMQVNLQINSKHNHSOKTSELLARTAYLGNKERCQL
SLNKIINQIVSFFYF"
98036. .98068
/gene="Cj1381"
/note="PS00013 prokaryotic membrane lipoprotein lipid
attachment site"
complement(98587. .99078)
/gene="fldA"
complement(98587. .99078)
/note="Cj1382c, fldA, flavodoxin, len: 163 aa; similar to
many e.g. FLAV_ECOLI flavodoxin 1 (175 aa), fasta scores;
opt: 405 z-score: 485.7 E(): 9.3e-20, 40.3% identity in
199 aa overlap. 41.1% identity to Hpl161. Contains PS00201
Flavodoxin signature, PS00017 ATP/GTP-binding site motif A
(P-loop), and Pfam match to entry PF00258 flavodoxin"
/codon_start=1
/transl_table=11
/product="flavodoxin"
/protein_id="CAB73809.1"
/db_xref="GI:6968816"
/translation="MSVAVIYGSAGMNTGAGNTIASKIGISDVFNISDIAAKMNSY
DKLIGTSTWGSGLQDDWDGDFSGLSLGGKTVAVFGNGDSESYSDTFCGGMGLAQ

NLKDAGANLVGVSTDGYTFEASDAVVDDGKFVGLALDNDNDQEDQTESRIDAWVEQIKP
YFA"
complement(98617. .99075)
/gene="fldA"
/note="Pfam match to entry PF00258 flavodoxin,
Flavodoxins, score 165.00, E-value 1.3e-45"
complement(98848. .98871)
/gene="fldA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(99016. .99066)
/gene="fldA"
/note="PS00201 Flavodoxin signature"
complement(99078. .99773)
/gene="Cj1383c"
complement(99078. .99773)
/gene="Cj1383c"
/note="Cj1383c, unknown, len: 231 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1383c"
/protein_id="CAB73810.1"
/db_xref="GI:6968817"
/translation="MLFGDDKQEFIPQIYRYLNNQELMLTQYNASVDSALKIPL
LYAKNTSLKMFIGNFLHMVSGKIQINIKLNTYAFYFQKRKSLJFNFKISKNV
DLRLRLIYLCIFCDAQILFSYVDKVFQNGKNIDQDGLIITIDKKFAILPLCK
ELNTHNLEIENCIYELLNLKENFEKFIYVCPRNKNFTFTEIKHFLCDLNKTMKLK
VPYKISNQIIRBK"
complement(99776. .10090)
/gene="Cj1384c"
complement(99776. .10090)
/note="Cj1384c"
/gene="Cj1384c, unknown, len: 104 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1384c"
/protein_id="CAB73811.1"
/db_xref="GI:6968818"
/translation="MSVLVIGADEITPIRAVLHDLGAKKIEHWDARNENVRNKPIPC
DTECIVMLTSFLNHTMKIKNEAKRKRIPLYCAKRSVSCYCEYCKIFNLNKEFSY
KG"
100261. .101685
/gene="kata"
100261. .101685
/gene="kata"
/EC_number="1.11.1.6"
/note="Cj1385, kata, catalase, len: 474 aa; 98.5%
identical to TR:Q59296 (EMBL:X85130) C. jejuni catalase
(BC 1.11.1.6) (507 aa), and similar to e.g. CATA_BRUAB
catalase (498 aa), fasta scores; opt: 2093 z-score: 2432.5
E(): 0. 65 6% identity in 474 aa overlap. 56.1% identity
to HP0875. Contains PS00437 Catalase proximal heme-ligand
signature, and Pfam match to entry PF00199 catalase"
/codon_start=1
/transl_table=11
/product="catalase"
/protein_id="CAB73812.1"
/db_xref="GI:6968819"
/translation="MKKLTNDFGNIIADNQNSLSAGAKPLLMQDYLLLEKLAHONRE
RIPERTVHAKSGAYGEKITADLSAYTKAKIFQGEVTPLEFRSTVAGEAADA
FWSLCPESLHOVTILMSDRGIPASVYRMHMGFSGHTYPSFINDKNERFWKHFKEKTOGI
KNLTNQEAEELAKDRESHORDLYNAIKNDOPPKWQVOIILAEKDIEKLGPNPDILT
KIWHSFVPLMDIGEMILNKKPNQYFNVEQAAFSNIVFGISPKMLQARIPSY
PDAQIRGTYNTHLLPVRNRAKSEVNTVAGAFNFSKNDAAAYEPSNIDYNSPKEDK
SYLEDPLVLEGVQARYAPLDNDFTQPRALFNLMDNDOKTQLFHNIASMEGVDEKII
TRALKHFEKISPDYAKGKKALEK"
100273. .101679
/gene="kata"
/note="Pfam match to entry PF00199 catalase, Catalase,
score 960.50, E-value 4.3e-285"
101248. .101274
/gene="kata"
/note="PS00437 Catalase proximal heme-ligand signature"

```

/db_xref="GI:6968808"
/translation="MKIIATSTNKHKVLFLKELKDFELIYAEDEVLMPEIEBNGKTF
KENALIKARAVNALDEKOKDFIALSDSGICVDLEGNPGIYSAREFGKDDKSNR
DKLVNEMIKRGKQSKAYVAALAMVAGMLGEFTHTGTMKGVIDTEKENGFGYDLSF
IPKGFDTLALQISVDEKNKNSHRFALDELAKIILKMGK"
/gene="91585"
/gene="Cj1375"
90290..91585
/gene="Cj1375"
90290..91585
/gene="Cj1375"
90290..91585
/feature="Cj1375"
/note="Cj1375, probable efflux protein, len: 431 aa;
similar to members of the major facilitator family e.g.
BMR1, BACSU multidrug resistance protein 1
(multidrug-efflux transporter 1) (389 aa) fasta scores;
opt: 364 z-score: 415.1 E(): 8e-16, 27.1% identity in 384
aa overlap, and NORA-STAU quinolone resistance NORA
protein (388 aa), fasta scores; opt: 353 z-score: 402.8
E(): 3.9e-15, 26.7% identity in 375 aa overlap. 41.9%
identity to Hpl181. Also similar to Cj0035c (24.7%
identity in 372 aa overlap). Contains pfam match to entry
PF00083 sugar_tr, Sugar (and other) transporters"
/codon_start=1
/transl_table=11
/product="putative efflux protein"
/protein_id="CAB73802.1"
/db_xref="GI:6968809"
/translation="MLNNVLPISFVIGTRFEGFLVLYSLYALKGANEFLVGL
VGVYALTMIIOMPFGIISDKTGRKKTMLIGITIFSLISCEFAENITMILGMLQ
GAGAIQAVATWISDFITEENKGMAMVSGFISGFASWISPLMSKMGSLSLFD
LSAALSLICILLYTVVPEKNKITHENKTPFPHIKNLALNLTJNMOMLSIA
FLSIPIILVHKGFSADKLVIYASVAVGAFAMGASLGKGLAKOAILLIGVVF
ILSYLFFAPNSIDPEITIAVVFYGFPMHEPIOMGASCKCKYHEKKAALGLPFAFG
YGGSPFGLIGIGIFLHDLTILNLAIIYLAIVYALFPLKNPDRFKNLVPLEPL
NPSFGENLGVYDIIKRNKNLVKFSKLSKELEGK"
90311..91432
/misc-feature
/gene="Cj1375"
/note="Pfam match to entry PF00083 sugar_tr, Sugar (and
other) transporters"
91582..92391
/gene="Cj1376"
91582..92391
/gene="Cj1376"
91582..92391
/note="Cj1376, probable periplasmic protein, len: 269 aa;
similar to TR:032102 (EMBL:299120) B. subtilis YU11
protein (314 aa), fasta scores; opt: 209 z-score: 257.4
E(): 4.9e-07, 30.6% identity in 242 aa overlap, and
TR:054506 (EMBL:097227) Salmonella enterica IROE (305 aa),
fasta scores; opt: 195 z-score: 240.9 E(): 4.1e-06, 26.4%
identity in 277 aa overlap. No Hp match. Contains probable
N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73803.1"
/db_xref="GI:6968810"
/translation="MKILFLFCSPFLAFAPKSLIEPLSKOAKONFLTEWKNYNG
TIYKIFIAKNKNSNYKATYIYLLDGNAPPLPLNLTNTYKKNLITGCIYNNLA
PDQKRTKRDYPTNTNKKQGGSKFRNRLTLQPLPYTKMIPIDFSQJLFGSFGG
LPALDITLDISKIFSHYFIISPDLMDNSFLPNIKLQTCPKIYIGIGSETSSII
KGVAKLNAOKLSEKISRETCKTHFKLEFKNETHGSVSKAMLMKMLDLN"
complement(92384..94045)
/gene="Cj1377c"
complement(92384..94045)
/gene="Cj1377c"
94045..94045
/note="Cj1377c, probable ferredoxin, len: 553 aa; limited
similarity to e.g. HYCF_ECOLI formate hydrogenlyase
subunit 6 (180 aa), fasta scores; opt: 212 z-score: 241.0
E(): 4e-06, 30.6% identity in 134 aa overlap. No Hp match.
Contains PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature, and 2x Pfam match to entry PF00037 fer4,
4Fe-4S ferredoxins and related iron-sulfur cluster binding
domains"
/codon_start=1
/transl_table=11

```

```

/product="putative ferredoxin"
/protein_id="CAB73804.1"
/db_xref="GI:6968811"
/translation="MODFVIRKNDVLIPLDAIEILDQANDKEALICNDKNQAOIYA
PEINFLKNSODEIILQSKNNVLTIVARASVYDGLDDEDSKQVORNLVADDAQTV
ELKEHEFRVIALSKEIILAFVSGVELDAVANKQEEVEIDDFLLFAEDLSVVRK
DFTROGCVNLNLFENLEVLFELOSRSKRYKATISYNAVSCVQHERRSERACACA
ETCPVAIILKDDENKHLFEQSDCCGCGICVCSGSLDAPMPRESFTCEPFYKD
KKILIPKKSLENLNLALPKDVLPMIGEMKLSMHELTLLONGKANLVETDVS
RGSNEIIVLNLNFEFERKPOKKAIPVAKDEKELONYLKPEFIODLKADYLNNTLTRE
NFAKRNQEKIKEDRESMSGEMLRGKKEINPTNTCLSCVGCNVALADQEN
AKFNASLDTTGTCYELSCAEKDTLKFPSGMEFASVFEYQTMKDELFPACIEGKE
FAIKRAVERIALMLKRFENDNKTITLYCCADCKAKVIMKAMIN"
complement(92639..92785)
/gene="Cj1377c"
92785..92785
/note="Pfam match to entry PF00037 fer4, 4Fe-4S
ferredoxins and related iron-sulfur cluster binding
domains. score 24.00, E-value 1.1e-05"
complement(93266..93409)
/gene="Cj1377c"
93409..93409
/note="Pfam match to entry PF00037 fer4, 4Fe-4S
ferredoxins and related iron-sulfur cluster binding
domains. score 28.70, E-value 4.7e-07"
complement(93287..93322)
/gene="Cj1377c"
93322..93322
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature"
94123..95445
/gene="Cj1377c"
95445..95445
/gene="selA"
94123..95445
/gene="selA"
94123..95445
/feature="selA"
/ec_number="2.9.1.1"
/note="Cj1378, selA, probable L-seryl-tRNA(Sec) selenium
transferase (selenocysteine synthase), len: 440 aa;
similar to e.g. SELA_ECOLI L-seryl-tRNA(Sec) selenium
transferase (EC 2.9.1.1) (463 aa), fasta scores; opt: 781
z-score: 867.6 E(): 0.37.5% identity in 451 aa overlap.
32.8% identity to Hpl513"
/codon_start=1
/transl_table=11
/product="L-seryl-tRNA(Sec) selenium transferase"
/protein_id="CAB73805.1"
/db_xref="GI:6968812"
/translation="MNRKTRFPQINTLIEDELSKSPYIRAFCKKVVAKLKNFQ
DEISKDLLELKEIKETFRYKDLQSVINASGVVHTNMGVSHHEELYEACQKICIN
YSNVEFDLNGRSGRYALVLEKMLPECEGALVNNNAAVPLVLSICYDKRQVIS
SRGEIVETGSRPVEVYKACGVKICEGTSKTKLKYEOAINENTAILKTRKSNF
ALMGHSHSVNTRKDHLAKERKELSYDYGSGWCENLNEKLIKPNPKIRKLVQECDIL
SFGDKLFGSVQAGIILIGKKEILEKLNQNLRLMRLVDTLSPNESLAKAYLOKDYE
KITLKLNLNDLSFEKRALROKELKROTOLKSKSLVGGSGMPDKSLDTYIILTFQD
DALKLQTRFRKRENTIGRIENDEFVLDFRTIRENELKILITNQMENL"
95442..97247
/gene="selB"
97247..97247
/note="Cj1379, selB, probable selenocysteine-specific
elongation factor, len: 601 aa; similar to e.g. SELB_ECOLI
selenocysteine-specific elongation factor (614 aa), fasta
scores; opt: 479 z-score: 491.6 E(): 4.4e-20, 26.2%
identity in 362 aa overlap. No Hp ortholog. Contains
PS00017 ATP/GTP-binding site motif A (P-loop), and Pfam
match to entry PF00009 GTP_LFTU, Elongation factor Tu
family"
/codon_start=1
/transl_table=11
/product="putative selenocysteine-specific elongation
factor"
97247..97247
/protein_id="CAB73806.1"
97247..97247
/db_xref="GI:6968813"
/translation="MKSVITIGTAGHDGKTSILKALNGFPGSDLSKEQEROITINLS
FSNLKLDKNISFLDVGHKDLQVWVSGAFGSCVLEFVVDINGSLKQESLEHLIK
ILDIKNIILVSKCDLCEINQKQSVLELEELKNDPLKLVKHSINNOGIEELKIK
LYTLENKENDELEIFHYIIDVPSLGGVYVCSLNGSTLTIKTIKICDITQCELYI
KNIOHDNENQOIKACNRVALSLNCDVKRELKGLILSKGTFKFAKEDALVAKNIOG
NSKMIFCVGRQIECKINILKLENDEFVHFSPDKNVFLSFDEAFILLONNVIGGC

```

/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1368"
/protein_id="CAB73795.1"
/db_xref="GI:6968802"
/translation="MKNLITKLENERLNOEENALMDLDTLTGKVAHKRRTLHGK
KVFENINRHNINPTNACDCKFCAPSARHKNNPYLMTHEELMKIVDEIVKKEVEVH
IVSAHNKDTQWWELEIFRMKEKYPNLMKMTAAEIDFLRRRMSYEVEIEMLE
YGVDSMPGGAEIFDEEVRKIKCHGKVSNNLKTHKLHESKOSNATMTLEGHIEER
HHRIDHMLRLNDLOGTGPNAFIPLVMOROKSFOTQDKMSSEILITIASRLVID
NIKNIKAWRTVTLNLAMVAFGANDLDGTIEKESIOSAGAKSAGKSTLKTPTDML
KTSNLIPVEROSLDELNLTXY"
83921..85240
/gene="Cj1369"
83921..85240
/gene="Cj1369"
/note="Cj1369, probable transmembrane transport protein,
len: 439 aa; similar to many hypothetical membrane
proteins e.g. YJCD_ECOLI (449 aa), fasta scores: opt: 897
z-score: 1039.1 E(): 0, 34.7% identity in 429 aa overlap,
50.7% identity to Hpi175. Contains Pfam match to entry
PF00860 xan_ur-permease, xanthine/uracil permeases family"
/codon_start=1
/transl_table=11
/product="putative transmembrane transport protein"
/protein_id="CAB73796.1"
/db_xref="GI:6968803"
/translation="MSPEFKENNTSEKTEVYAGITFTFLAMYIILPVNSSTVGNMGMP
IEALTATALLITVASAFNAFFANTPVPAMSVGMGNATFTTAVCLGQNIAMOSALGAV
FISITIFILSTFTFRILWIRNIPKDLALICAGICFIAFLGLSOMGVNHNKQTLV
SIGNKSPHVEIGAEITLLIIFEMAIKRGAFILCLVASIIAMFIHDSNAPFOIF
SWPNSMENGLGAEIFLQIDIKSALNITMIPILTFITQLEPSITIGVSGRFIFD
DPKNEKLSKTLMDATGSAAGMTGSTVTAPESTIGVESGGRTGTLTALVAVICF
AFTLELPLKFAIPANAIYPVLVWVGITLMFMEVKKIDFKDSIAVASFTIIMMFTY
SITTFAPGFEISYLLVIRFKRMKINIGIIVLSLISGNFLMALQ"
83960..85120
/gene="Cj1370"
/note="Pfam match to entry PF00860 xan_ur-permease,
xanthine/uracil permeases family"
85254..85697
/gene="Cj1370"
85254..85697
/gene="Cj1370"
/note="Cj1370, possible nucleotide
phosphoribosyltransferase, len: 147 aa; similar to e.g.
TR-050284 (EMBL:AF035967) hypoxanthine ribosyl transferase
(176 aa), fasta scores: opt: 158 z-score: 205.5 E():
0.00038, 26.2% identity in 145 aa overlap, and XGPT_ECOLI
xanthine-guanine phosphoribosyltransferase (152 aa), fasta
scores: opt: 128 z-score: 170.3 E(): 0.035, 24.6% identity
in 142 aa overlap, 39.5% identity to HP0735. Contains
PS00103 Purine/pyrimidine phosphoribosyl transferases
signature"
/codon_start=1
/transl_table=11
/product="putative nucleotide phosphoribosyltransferase"
/protein_id="CAB73797.1"
/db_xref="GI:6968804"
/translation="WIFYSEEFKEDVKILAEKEDFNPDALLAIANGSLGHSIA
VALKTRQLEFALNSIHYDNTKRLDTIEINIDLSHKKLTLLIDVDGSESLAEIKV
LLEKFPHEIKTITFTFYKTKALLBPDKFKATEWVDYMDIND"
85500..85538
/gene="Cj1370"
/note="PS00103 Purine/pyrimidine phosphoribosyl
transferases signature"
85742..86440
/gene="Cj1371"
85742..86440
/gene="Cj1371"
/note="Cj1371, probable periplasmic protein (vacJ
homolog), len: 232 aa; similar to e.g. VACJ_SHIFL VACJ
lipoprotein precursor (required for intercellular
spreading of Shigella flexneri) (251 aa), fasta scores:

opt: 310 z-score: 377.2 E(): 1e-13, 32.5% identity in 197
aa overlap. No Hp match. Contains probable N-terminal
signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein (vacJ homolog)"
/protein_id="CAB73798.1"
/db_xref="GI:6968805"
/translation="MRKIFISPLIFPTIIPAFKEDONDDPEOEYONYOVNDPLSGYN
KAMTSFNVALDXGILPVLKLPALPEIRLIGANFENDLLAPRPSVNTQKPEE
AGEEKRRTANTIKGFGLMDVASMGLKKRPADIGTLAHGVSCHRIYVLPILGPS
NLRDITLPAIWAYSAFTYVIDPTWASIAISAVGFENELSRUDEIDETIHNPNLPLP
LRDAYEORRNELSK"
86450..87019
/gene="Cj1372"
86450..87019
/gene="Cj1372"
/note="Cj1372, probable periplasmic protein, len: 189 aa;
similar to e.g. YRBC_ECOLI (211 aa), fasta scores: opt:
189 z-score: 222.3 E(): 4.4e-05, 22.5% identity in 200 aa
overlap. No Hp match. Contains probable N-terminal signal
sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAB73799.1"
/db_xref="GI:6968806"
/translation="MKKIFILITLFPNPAQLQDEISSTMOKNIDASIKLQDNKEK
KQANGITRLPDSITIDYKLMKLSLSKNTSKITPEQKFTTAEFTSLKSTDLKSL
YKQDYLKRNKNEBKNEKRPFTTSMVVGEEKNIIKFYNDNNMLIYDVLVGSIV
QTYRSQFDIILANQCFDALLQKLESIVIE"
87019..89490
/gene="Cj1373"
87019..89490
/gene="Cj1373"
/note="Cj1373, probable integral membrane protein, len:
823 aa; similar to hypothetical membrane proteins e.g.
TR-051268 (EMBL:AE001135) Borrelia burgdorferi conserved
hypothetical integral membrane protein (767 aa), fasta
scores: opt: 331 z-score: 348.2 E(): 4.2e-12, 24.8%
identity in 840 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73800.1"
/db_xref="GI:6968807"
/translation="WEAKLKFETISRPKSTFFGTLFTCLFSEFAFKSVDSASBSLL
LEDADLKFREISIKHYKSDNFLLAFKPYDEKPPSNNELAKKRLHBELEKAPVER
VFSIINADLQSSQNTDLKEELKNIPNIESODINRTKQNETLNSPEYKNNITSDKG
VTGLITVLPPTVYNNLEKRDLATDEKEQIRLAIEHODKOKVITKOSIDITKSI
VRNENKNDTLYIGGVSMIADDMIAVITKSDLVIGVSLVFLGLALYVFPSSWRPVL
PLFICISLSAAGVFPALNQIYIYSNVALVILITLSVYVHITFTISTQNHK
AKVERIVLETLAKANPSLVAIVTTWIGFSLISNIEPIILKGIWMSIGLALFIS
YLFLASILVLLPKPKYHKKERFENLACAKATADLPKRMATIESVATILALIGIS
KLRENSFVNYFKDSEIKKGLIVDKMLGCTLPLEVIRFPNNKNDONASDTLDSEF
SEFENLATQETVYFQPSKTRIAKVEHLEKKEPGLSVLSNLSTLTKNDNGEELD
DFALAFLENENPARKKODLISPEVSIENNELRESMRIYDSDPNLRNELLIDIKOLN
ELVKNQGEVOITGIMVLVYNNMLOSLSRSQDTLFLVLAIFLFIYFRLKLSVAA
ILVAVTLPISVVALMGILGIPDMWSITIAIAGIGVDAIHVIYFREREIKNSLE
EAIMSHISLSEALYVTTISIVIGSVSMSSNFIPTIYFGLITVFMILLSGSLFL
PSFLITVYKSKTEPQNHKEHILKQ"
/complement(89521..90123)
/gene="Cj1374c"
/complement(89521..90123)
/gene="Cj1374c"
/note="Cj1374c, unknown, len: 200 aa; similar to
hypothetical proteins e.g. TR:P94558 B. subtilis YSNA (198
aa), fasta scores: opt: 451 z-score: 528.4 E(): 3.9e-22,
42.4% identity in 198 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1374c"
/protein_id="CAB73801.1"


```

misc_feature complement(74003. .75340)
/gene="func"
/note="Pfam match to entry PF00206 lyase_1, Lyases, score
689.20, E-value 2e-203"
complement(74396. .74425)
/gene="func"
/note="PS00163 Fumarate lyases signature"
complement(75426. .78551)
/gene="Cj1365c"
complement(75426. .78551)
/gene="Cj1365c"
/EC_number="3.4.21.1"
/note="Cj1365c, probable secreted serine protease, len:
1041 aa; similar to e.g. TR:Q54484 (EMBL:D78380) Seriatia
marcescens serine protease homologue SSP-H2 (1034 aa),
fasta scores; opt: 245 z-score: 257.1 E(): 5.1e-07, 23.5%
identity in 591 aa overlap, and PRS:SERMA extracellular
serine protease precursor (EC 3.4.21.-) (1045 aa) fasta
scores; opt: 205 z-score: 214.6 E(): 0.00012, 23.7%
identity in 615 aa overlap. No hp match. Contains PS00137
Serine proteases, subtilase family, histidine active site,
PS00138 Serine proteases, subtilase family, serine active
site, Pfam match to entry PF00082 subtilase, Subtilase
family of serine proteases, and probable N-terminal signal
sequence"
/codon_start=1
/transl_table=11
/product="putative secreted serine protease"
/protein_id="CAB73792.1"
/db_xref="GI:6968799"
/translation="MKKFECLTVCKLALSEFELHIDKVKLGYSDDTIIIGVAD
APNODHSLKRLKSTPTAGKQLPDLKSTHSHVAGIAGAKIGSKPQVGA
YGAFFYGVGVPPNSVYQIPDIYNEFKDVSIIINNSGINTFYPLMKASNGLDVCTO
TNGSTVNICMTPLEYVKAKVANMDKSLKGVNLVPAAGNEGILSPALHILPS
YDESLRALVAGALDANEITLESDDTLIKSGDLDFSGKGAFTSLVAGAVIIN
VDSSTNDKFKKSGTSMAPVSGTAAIVKONPELDGKOADILLSTANAKVAKRE
TKVOTDGTNPKPLIYISODPRGIDPEIKRDLKQIINGIQVOYNGWIDYSITWD
NRDSQOSKLTWTSTISSINGVYRVEKEELFEQGIIDAKAKGSLIDANRLSDDDL
KYDEPNATVYTTINAGYDAEFNSNDISQKWDSETHLSAINKPHLANLNGISKEG
EGIIISGNTVEGATLTKOGLKLGKNNAYEOKAISLNGIYVONINNKIKYR
PGNEIDLNTQVGTQYGVDSKQLDQGNKNSKLAKTYDIKSGNLEYIPLEYYI
LNKPKYKILGDLKSLSEFNHVLIONYFALNDPVLSDILVSIKNTLILKPKMLPAYE
IPNTSIGNALROLRSRADLSQYQDFASLDNGIDVKTAKIRBSGSLSPSNHNS
NLQONNMFLTHPLINNFQONNNTLASTVLPRIFSNEEYFHWLTPSYKYKDDFS
GQKTGANSTLSENFSSGLAYALSSSAKFNPNNSDLKSYNMDLLVNYHDLDFIKI
LSGLGIGVGFNLNRFVEQPIEGEYKTLQTSAGQVYKDIILGQDFENFLMFTTH
FFYQDEFEKNSPPAKNYESLKHSHINANLGNFLAKNIEDDYQOASTFYIFEKRIY
GRTLENKASFVDPPLAFIQYKKKLNILISQGFNSFEFLKNNVFQFMILNMFSHAYE
LHLMSSVGRK"
complement(77511. .77972)
/gene="Cj1365c"
/note="Pfam match to entry PF00082 subtilase, Subtilase
family of serine proteases, score 31.80, E-value 8.5e-08"
complement(77652. .77684)
/gene="Cj1365c"
/note="PS00138 Serine proteases, subtilase family, serine
active site"
complement(78225. .78485)
/gene="Cj1365c"
/note="Pfam match to entry PF00082 subtilase, Subtilase
family of serine proteases, score 4.80, E-value 2.5"
complement(78282. .78314)
/gene="Cj1365c"
/note="PS00137 Serine proteases, subtilase family,
histidine active site"
complement(78554. .80350)
/gene="gIms"
complement(78554. .80350)
/gene="gIms"
/EC_number="2.6.1.16"
/note="Cj1366c, gIms, probable
glucosamine-6-phosphate aminotransferase
(isomerizing), len: 598 aa; similar to e.g. GLMS_ECOLI

```

```

misc_feature complement(79808. .80347)
/gene="gIms"
/note="Pfam match to entry PF00310 GATase_2, Glutamine
amidotransferases class-II, score 222.20, E-value 4.3e-75"
complement(80355. .82703)
/gene="Cj1367c"
complement(80355. .82703)
/gene="Cj1367c"
/note="Cj1367c, possible nucleotidyltransferase, len: 782
aa; weak similarity to TR:Q53245 (EMBL:U07030) Rhizobium
tropical putative uridylyltransferase/uridylyl-removing
enzyme (971 aa) fasta scores; opt: 202 z-score: 226.1
E(): 2.7e-05, 19.7% identity in 85 aa overlap, and
GLND_HAEIN [protein-pil] uridylyltransferase (863 aa),
fasta scores; opt: 119 z-score: 133.2 E(): 4.19e-5%
identity in 822 aa overlap, (blastp score: E-4.3e-07). No
hp match. Contains C(1-2) polymorphic region at aa 148:
C(2) (consensus) gives this ORF, C(1) gives truncation
after 18 aa"
/codon_start=1
/transl_table=11
/product="possible nucleotidyltransferase"
/protein_id="CAB73794.1"
/db_xref="GI:6968801"
/translation="MLKKELEPFKKNLSSTYQSCRKFFLKQAGATSLYKMWDPFIK
AYIDLKDYFDDFSPSDNIPFCVLAASAAVANNSCFNESTISLJFYKVDIAFHLKPM
IKAFTEILDNAHLQIDSVILEFNGLYNANSBLKTSIIOTRFICGRSLIFKGIKIPES
IIKENKNDPFAKLLENPEKEDIPFIKQEPNLKDFGGLNHRSLSESLLVFKTSPKNY
ALNFIDEKKLSSELRLAGDFELSLKASAMULLSAKDEDLILNIVDLSLAWKRRKKR
GANELLVOKALQOSMHTGTFYHFLAKOQDOLNLTQKOEYKFKTVLVETVELLAKLEK
HYIDPLNVPALGRKLRKIDIEKALILFEKIFIKRHSFCVALLDLGSLDKLCKPR
WVRFPLDSEGNYSFDEQVFLMSEFEYEDDELQLOKTKDEKMTLKVILLASIES
ENNESLQAVIYANYSCKPDLKNEILEMGLIKFNNAALDVLDEKEDIVPIVSSLYSK
LENLENLELTYTLTWLAKALNYAVAFRVLDLLENNKOGFEDENILLESARVYKE
LTLKRSKIFLEODELQDKIITHIKSNLFIINTFEDIYMIKILAKENDKRFWSPEN
LSLOIVAPLHNHNAIILISLNLNLIJNPFELDDKRYLREEDYDNTISDQOKLCE
LLNLSLSEFNKIKRKPITRKDELKDLNTSKMKAKGLNTRKQOGLMAYLMMVFNEL
ELVLCARQITQTRKTRNIFIFQKNEKHESEKLVNLLISE"
82260. .82261
/note="polymorphism C(1-2)"
82847. .83911
/gene="Cj1368"
82847. .83911
/gene="Cj1368"
/note="Cj1368, unknown, len: 354 aa; similar to
hypothetical proteins e.g. TR:066888 (EMBL:AE000700)
Aquifex aeolicus AQ_648 (371 aa), fasta scores; opt: 956
z-score: 1117.5 E(): 0, 44.1% identity in 349 aa overlap.
57.4% identity to HP0654. Also similar to Cj0462 (35.5%
identity in 349 aa overlap)"

```



```

PF00005 ABC_tran, ABC transporters"
/codon_start=1
/transl_table=11
/product="enterochelin uptake ATP-binding protein"
/protein_id="CAB73781.1"
/db_xref="GI:696878"
/translation="MIKLKNTKPYDNKVIISDLSDFHKKITSITIGANGAKSTLL
ALASRLKIPSSGEIYIDGMNLKNYKQILAOIKSITLKOONNINLKAEEIYAPRFP
HSOGRLLINDKIKINEALEYELSNLRNPELDTSGGOKRAFIMITIAQDEFLMFD
EPINLIDMKHSVIMOLMKNLVDPKNSIAYVLIHINRASIYSDEILAKDKGLKOG
SKDEIINENLKOITDMDIPVSIQNGKICIFY"
63717..64274
/misc_feature
/ gene="ceud"
/ note="Pfam match to entry PF00005 ABC_tran, ABC
transporters, score 170.60, E-value 2.6e-47"
63738..63761
/misc_feature
/ gene="ceud"
/ note="PS00017 ATP/GTP-binding site motif A (P-loop)"
64044..64088
/ gene="ceud"
/ note="PS00211 ABC transporters family signature"
64407..65399
/ gene="ceud"
/ note="Cj1355, ceud, probable enterochelin uptake
periplasmic binding protein, len: 330 aa, highly similar
to TR:Q46006 (EMBL:X88849) Campylobacter coli ceud (328
aa), fasta scores: opt: 1903 z-score: 2111.3 E(): 0, 90.6%
identity in 330 aa overlap, and similar to e.g. FARB_VIBAN
feric anguibactin-binding protein precursor (332 aa),
fasta scores: opt: 540 z-score: 605.4 E(): 2e-26, 30.1%
identity in 279 aa overlap. No Hp match. Contains probable
N-terminal signal sequence and appropriately positioned
PS00013 Prokaryotic membrane lipoprotein lipid attachment
site"
/codon_start=1
/transl_table=11
/product="enterochelin uptake periplasmic binding protein"
/protein_id="CAB73782.1"
/db_xref="GI:6968789"
/translation="MKKSLVFAFAFPLSLITLACNSNSNENNSSTTKNTATVKYL
PISSDEGDSFLVDSLGENTIKPNKPSKVYLLDITPDALKNDKVVGVPAKNPL
KYLOQFNKRPVGVGVQVDFEAINALKPDILITIGROSKFEDKKEIAPTFVGLDIA
NFISSFNNVSVAKIYGLKEALEKISDINEKEAKSIYDEKKAIIITNSKIS
AFGQSRGIIHDVYGINAVDENTKVGHSINSEFLLENPDYIFVDNVTIENK
ERAQGLIDNALVAKTKAQNKKIITLDPYWYIASGNIESTLKITLEIKNAV"
64437..64469
/misc_feature
/ gene="ceud"
/ note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(63472..65561)
/ gene="tRNA-Ser"
/ note="tRNA Ser anticodon GCT, Cove score 68.31"
complement(65472..65561)
/ gene="tRNA-Ser"
complement(65625..66179)
/ gene="Cj1356c"
complement(65625..66179)
/ gene="Cj1356c"
complement(65625..66179)
/ note="Cj1356c, probable integral membrane protein, len:
184 aa: highly similar to TR:Q46007 (EMBL:X88849)
Campylobacter coli ORF (210 aa), fasta scores: opt: 545
z-score: 652.8 E(): 4.6e-29, 67.1% identity in 210 aa
overlap. 31.8% identity to Hp0234"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73783.1"
/db_xref="GI:6968790"
/translation="MINKNKLFRQVNIYLSLFLPCALLFALGIAVIFGINDVGILK
VEQYOLSKVIESGKEALIEYKTLNGLKVPMSNITIKSDKGLITIGTGYSANITON
STNEYNTLTLSLIDMIMLHKDKGAWPVSIVSGCITLFLMILYISGLMITTEAFANK
PF00005 ABC_tran, ABC transporters"
/codon_start=1
/transl_table=11
/product="enterochelin uptake ATP-binding protein"
/protein_id="CAB73781.1"
/db_xref="GI:696878"
/translation="MIKLKNTKPYDNKVIISDLSDFHKKITSITIGANGAKSTLL
ALASRLKIPSSGEIYIDGMNLKNYKQILAOIKSITLKOONNINLKAEEIYAPRFP
HSOGRLLINDKIKINEALEYELSNLRNPELDTSGGOKRAFIMITIAQDEFLMFD
EPINLIDMKHSVIMOLMKNLVDPKNSIAYVLIHINRASIYSDEILAKDKGLKOG
SKDEIINENLKOITDMDIPVSIQNGKICIFY"
63717..64274
/misc_feature
/ gene="ceud"
/ note="Pfam match to entry PF00005 ABC_tran, ABC
transporters, score 170.60, E-value 2.6e-47"
63738..63761
/misc_feature
/ gene="ceud"
/ note="PS00017 ATP/GTP-binding site motif A (P-loop)"
64044..64088
/ gene="ceud"
/ note="PS00211 ABC transporters family signature"
64407..65399
/ gene="ceud"
/ note="Cj1355, ceud, probable enterochelin uptake
periplasmic binding protein, len: 330 aa, highly similar
to TR:Q46006 (EMBL:X88849) Campylobacter coli ceud (328
aa), fasta scores: opt: 1903 z-score: 2111.3 E(): 0, 90.6%
identity in 330 aa overlap, and similar to e.g. FARB_VIBAN
feric anguibactin-binding protein precursor (332 aa),
fasta scores: opt: 540 z-score: 605.4 E(): 2e-26, 30.1%
identity in 279 aa overlap. No Hp match. Contains probable
N-terminal signal sequence and appropriately positioned
PS00013 Prokaryotic membrane lipoprotein lipid attachment
site"
/codon_start=1
/transl_table=11
/product="enterochelin uptake periplasmic binding protein"
/protein_id="CAB73782.1"
/db_xref="GI:6968789"
/translation="MKKSLVFAFAFPLSLITLACNSNSNENNSSTTKNTATVKYL
PISSDEGDSFLVDSLGENTIKPNKPSKVYLLDITPDALKNDKVVGVPAKNPL
KYLOQFNKRPVGVGVQVDFEAINALKPDILITIGROSKFEDKKEIAPTFVGLDIA
NFISSFNNVSVAKIYGLKEALEKISDINEKEAKSIYDEKKAIIITNSKIS
AFGQSRGIIHDVYGINAVDENTKVGHSINSEFLLENPDYIFVDNVTIENK
ERAQGLIDNALVAKTKAQNKKIITLDPYWYIASGNIESTLKITLEIKNAV"
64437..64469
/misc_feature
/ gene="ceud"
/ note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(63472..65561)
/ gene="tRNA-Ser"
/ note="tRNA Ser anticodon GCT, Cove score 68.31"
complement(65472..65561)
/ gene="tRNA-Ser"
complement(65625..66179)
/ gene="Cj1356c"
complement(65625..66179)
/ gene="Cj1356c"
complement(65625..66179)
/ note="Cj1356c, probable integral membrane protein, len:
184 aa: highly similar to TR:Q46007 (EMBL:X88849)
Campylobacter coli ORF (210 aa), fasta scores: opt: 545
z-score: 652.8 E(): 4.6e-29, 67.1% identity in 210 aa
overlap. 31.8% identity to Hp0234"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB73783.1"
/db_xref="GI:6968790"
/translation="MINKNKLFRQVNIYLSLFLPCALLFALGIAVIFGINDVGILK
VEQYOLSKVIESGKEALIEYKTLNGLKVPMSNITIKSDKGLITIGTGYSANITON
STNEYNTLTLSLIDMIMLHKDKGAWPVSIVSGCITLFLMILYISGLMITTEAFANK
DRSKQFAVLGVGVVYPLLAVLSL"
complement(66315..68147)
/ gene="Cj1357c"
complement(66315..68147)
/ gene="Cj1357c"
complement(66315..68147)
/ note="Cj1357c, probable periplasmic cytochrome C, len:
610 aa: similar to e.g. NRFL_ECOLI cytochrome C52
precursor (478 aa), fasta scores: opt: 298 z-score: 333.9
E(): 2.7e-11, 31.0% identity in 455 aa overlap. No Hp
match. Contains probable N-terminal signal sequence and 5x
PS00190 Cytochrome c family heme-binding site signature"
/codon_start=1
/transl_table=11
/product="putative periplasmic cytochrome C"
/protein_id="CAB73784.1"
/db_xref="GI:6968791"
/translation="MKKNILRIGIVVILLAGVILMNNDINOKKDEANKRAITANA
DESLIDDDPNFEKKGKVPKPOLKMYLVREKPEKATPEGGLATSKILRPOLITLW
AGYPSLDPNERGFRFVYVDKMTAKRNKDFLNHGLAARFGOPACACNSGWTW
LTKNVAKGDFTAFNSTNWTMKNIPAVDGIENSPEHAGPHGKRMGVTCADCHNP
DMSLRLTPRAAINALVSRYGDEYFPMHDKNKVEITFDGIVLFPSEMKKGOPRI
MGETIVDSSKKNMGTOKNYDEYFPMHDKNKVEITFDGIVLFPSEMKKGOPRI
EMLDYVDRKVGVEGADFTHLKLGQIITKIOHPSELYSGGVHAANGVCYDCHNPV
REGAKKVTQHNITSPLRDINSACKSCHQSEDIYLAQVYLDIONSVAHQRTAETAVS
LIMDTKRLDELGNEKFOSDGKADAKTISELKEVLELHRRADQRAQVAENSTGF
HNPRASRLQAVDMAMRGOTKLEIILAAANGIDFKTSNLGFPEDIOKFNPEGLYKV
DVNNIKAGEERYADKEDVNGPKLELHDELAAYNYVIDK"
complement(66888..66905)
/ gene="Cj1357c"
complement(66888..66905)
/ note="PS00190 Cytochrome c family heme-binding site
signature"
complement(66981..66998)
/ gene="Cj1357c"
complement(66981..66998)
/ note="PS00190 Cytochrome c family heme-binding site
signature"
complement(67359..67376)
/ gene="Cj1357c"
complement(67359..67376)
/ note="PS00190 Cytochrome c family heme-binding site
signature"
complement(67500..67517)
/ gene="Cj1357c"
complement(67500..67517)
/ note="PS00190 Cytochrome c family heme-binding site
signature"
complement(67683..67700)
/ gene="Cj1357c"
complement(67683..67700)
/ note="PS00190 Cytochrome c family heme-binding site
signature"
complement(68162..68677)
/ gene="Cj1358c"
complement(68162..68677)
/ gene="Cj1358c"
complement(68162..68677)
/ note="Cj1358c, probable periplasmic cytochrome C, len:
171 aa: similar to e.g. NARC_ECOLI cytochrome C-type
protein NARC (200 aa), fasta scores: opt: 201 z-score:
242.9 E(): 3.1e-06, 34.5% identity in 165 aa overlap. No
Hp match. Contains probable N-terminal signal sequence and
3x PS00190 Cytochrome c family heme-binding site
signature"
/codon_start=1
/transl_table=11
/product="putative periplasmic cytochrome C"
/protein_id="CAB73785.1"
/db_xref="GI:6968792"
/translation="MILRRKILKTSNLPTIFVILFPAVGFYFNNAKGTSYLS
NASSCSNCHIMNEVYENYAGVSCVKEPRATCDVCLPHRFVAKKATKAGSGIG
HAYAFTRKDEBLPPLNSATKSRKMOVENCIRCHADFAQTAINATVPHADKSLNCAS
CHNDVGRKHGI"
complement(68189..68206)
/ gene="Cj1358c"
complement(68189..68206)
/ note="PS00190 Cytochrome c family heme-binding site
signature"
complement(68423..68440)
/ gene="Cj1358c"
complement(68423..68440)

```

```
/translation="MPVSPGLGINVYINONNAPATQASNELAKGEAATINMAEFENK
EKTINLEKNEVTHEIKEIKERKAEOEKKKHQAKTSDPEKNEDESKDPFKDA
QSIRHLDISI"
complement(58727..60034)
/gene="Cj1349c"
/codon_start=1
/translation="MKYTELLQLOTFPFRPKIDFIKRIINDNILESPDEREFIDLT
RAMSAIYIAKFNKANYNAPEDFMILKYSNAFIKEVYLEGNRILICFSVKANKKYSY
ESKIYEPFGKNTNVLITDERDLITELRHIDKSYRVKPNVLESILKPKMDENPEE
IKDEADYFSRKPTSYESKIKQIKNLKLAOYDKTIQNLQELFSSLDSENSILKALEY
RKRADVLPANLSVKDYERERKIDDEGKEDEKLEFNPKOSANLYIKNKKLEQKK
NLNIQRONLEKLDPELTLKELLOAKSELELILPKKSKKNOHKOEDLVANFYE
NEERTICVGNKNGNEFLKNKAKDDMLHVRDIPSAHTLIIVSNOKISLDVIEFARL
CVSFSLKSGSYWDYTLKNFVKQAKAFVNTNFKSINITKD"
60096..60671
/gene="mba"
60096..60671
/translation="Cj1350, moba, probable molybdopterin-guanine
dinucleotide biosynthesis protein, len: 191 aa; similar to
e.g. MOBA_ECOLI molybdopterin-guanine dinucleotide
biosynthesis protein A (194 aa), fasta scores: opt: 194
z-score: 249.0 E(): 1.5e-06, 22.9% identity in 192 aa
overlap, 26.6% identity to Hp0769"
/codon_start=1
/translation="putative molybdopterin-guanine dinucleotide
biosynthesis protein"
/product_id="CAB73777.1"
/protein_id="CAB73777.1"
/db_xref="GI:6968784"
/translation="MOLNELNCVILCGSKSRMGODSKLILKNQNLTOPQVDFSKI
FKVNYSAKEDKEENHSLIKSLSEFVSPMLATYISLNFKNKFEVYFVSVSPKG
ENELIKMLPELEONYKIIIAKTPLHKKPLCGFYHSSLAQCKNLEKNECKIGLSESE
IKTKFVEDEDDAFILNFEYEEKFSKSK"
60668..61657
/gene="pida"
60668..61657
/translation="Cj1351, pida, probable phospholipase A, len: 329
aa; highly similar to TR:032349 (EMBL:Y11031)
Campylobacter coli phospholipase A (297 aa), fasta scores:
opt: 1544 z-score: 1835.1 E(): 0.75.2% identity in 298 aa
overlap, and similar to e.g. PAI_ECOLI phospholipase A1
precursor (EC 3.1.1.32) (289 aa), fasta scores: opt: 467
z-score: 560.5 E(): 6.4e-24, 34.7% identity in 265 aa
overlap, 25.5% identity to Hp0499. Contains probable
N-terminal signal sequence"
/codon_start=1
/translation="phospholipase A"
/product_id="CAB73778.1"
/protein_id="CAB73778.1"
/db_xref="GI:6968785"
/translation="MRKIALFLSLCVFIWASDIQALAEYKQDYKKAMEYKRLAK
NSSVLISQEONNSSOATQTONSITIKKEKODFSRLALANLGENEFNGLISYKM
NYLFPAYSENLSGVNNKSEAKFOLSKRLEFNGLIDEXYIAYTOSMOQIYEH
SSPREFNVOPEFIDPLTYIDYEPFNRIAGVILHESNGKDENLOSBSMNLIYST
ALYKNTFLPYRIPRYRIPENKDDNDPALIHYMGNDPVNATYAGDYFTIMLRNNIK
FHNKGAIOVDGIDFNNGITWYLOTFNGISGLIDYKHLORLSTGFLIST"
61743..62711
/gene="ceub"
61743..62711
/translation="Cj1352, ceub, probable enterochelin uptake
permease, len: 322 aa; highly similar to TR:Q46003
(EMBL:X88849) Campylobacter coli ceub (322 aa), fasta
scores: opt: 1839 z-score: 2038.7 E(): 0.89.1% identity
in 322 aa overlap, and similar to e.g. FATD_VIBAN ferric
anquibactin transport system permease protein FATD (314
aa), fasta scores: opt: 921 z-score: 1056.8 E(): 0.43.3%
identity in 305 aa overlap. Also similar to Cj1615 (21.0%
identity in 326 aa overlap), and Cj1353 (27.2% identity in
316 aa overlap). Contains Pfam match to entry PF01032
FecCD_family, FecCD transport family"
/codon_start=1
/translation="enterochelin uptake permease"
/product_id="CAB73779.1"
/protein_id="CAB73779.1"
/db_xref="GI:6968786"
/translation="MPFKHLSLKLVALLLFGMISLFTGISVINKYDILNLSIOL
EILTRIPRIALILMSISIGLMOODTOKRFPPTAGMDCARGLISLIF
FGASPTFOALISIFALSGSFIOIRIKIKDIFVPLIGMEGGIISATTEFA
YALNTYIONIGMWSMANNQGNVELTISLPLIATYLAKITTYGGEIDALNU
GISTNGILFGLMVSITSLVTSVGIPLFLGILPVALYDGLNRNLNLYTALC
GALFLVCDIISRLVIFPEMPMLSTTGVGSLIFFLKRRVYA"
61821..62693
/gene="ceub"
61821..62693
/translation="Pfam match to entry PF01032 FecCD_family, FecCD
transport family, score 174.90, E-value 1.4e-48"
62704..63642
/gene="ceuc"
62704..63642
/translation="Cj1353, ceuc, probable enterochelin uptake
permease, len: 312 aa; highly similar to TR:Q46004
(EMBL:X88849) Campylobacter coli ceuc (312 aa), fasta
scores: opt: 1784 z-score: 2112.8 E(): 0.88.8% identity
in 312 aa overlap, and similar to e.g. FATC_VIBAN ferric
anquibactin transport system permease protein FATC (317
aa), fasta scores: opt: 593 z-score: 708.6 E(): 3.6e-32,
31.7% identity in 305 aa overlap. No Hp ortholog. Also
similar to e.g. Cj1352 (27.2% identity in 316 aa overlap),
and Cj1615 (27.5% identity in 298 aa overlap). Contains
Pfam match to entry PF01032 FecCD_family, FecCD transport
family"
/codon_start=1
/translation="enterochelin uptake permease"
/product_id="CAB73780.1"
/protein_id="CAB73780.1"
/db_xref="GI:6968787"
/translation="MRKKMLISFLTLNMGIFIPVGLNGFDEYALKSRFLQIAIILI
VAICIAVSIVYIQTLCNNKILTPAIIIGDSLWMLQSALEFSGANLSYKNDINPL
ITIVCAWVPSLGKIKLIFSSDRSRIYIMLGLIYFGLFSLSEFVYLLDPEFMTQ
GRMFASDNIAFDVLIDATITSLFTWRFKFDPDLGDLNINGLINQIKSK
QLMIIAIIITSLTALVGPITPLGLVNTTYELFTAKHSILLSACIISILALGG
VFPEVSRYFDNATISVINPLGIVFYLVKGNKL"
62728..63627
/gene="ceuc"
62728..63627
/translation="Pfam match to entry PF01032 FecCD_family, FecCD
transport family, score 64.10, E-value 3.1e-15"
63639..64394
/gene="ceud"
63639..64394
/translation="Cj1354, ceud, probable enterochelin uptake
ATP-binding protein, len: 251 aa; highly similar to
TR:Q46005 (EMBL:X88849) Campylobacter coli ceud (251 aa),
fasta scores: opt: 1258 z-score: 1347.9 E(): 0.82.5%
identity in 246 aa overlap, and similar to e.g. FHUC_BACSU
ferrichrome transport ATP-binding protein (269 aa), fasta
scores: opt: 613 z-score: 663.4 E(): 1.2e-29, 38.0%
identity in 245 aa overlap. No Hp ortholog. Contains
PS00017 ATP/GTP-binding site motif A (P-loop), PS00211 ABC
transporters family signature, and Pfam match to entry
```

```

misc_difference
53102..53110
/note="polymorphism C(9-10)"
gene
complement(53670..54185)
CDS
/ gene="Cj1343c"
complement(53670..54185)
/ note="Cj1343c"
/ gene="Cj1343c"
/ note="Cj1343c, probable periplasmic protein, len: 171 aa;
no Hp match. Some similarity to N-terminus of secretion
proteins e.g. GSPQ.ECOLI putative general secretion
pathway protein (145 aa), fasta scores: opt: 153 z-score:
192.8 E(): 0.0019, 35.6% identity in 59 aa overlap.
Contains probable N-terminal signal sequence"
/codon_start=1
/transl_table=1
/product="putative periplasmic protein"
/protein_id="CAB73770.1"
/db_xref="GI:6968777"
/translation="MKKAFITLLEVPVIVILGILAAIALPRMSSSKDEAEVSKINL
KTLINDISITYLKNDHLSIKTMNSVSGVENVDLSNFGIKENVRVGDDECKLTVF
INKADFLMGISSNEASKNAIINAQSHEDLENIDFTSSSNKACVILSKNENPKNL
ASKTYLLIGCM"
gene
complement(54186..55193)
CDS
/ gene="Cj1344c"
complement(54186..55193)
/ gene="Cj1344c"
/ note="Cj1344c, probable glycoprotease, len: 335 aa;
similar to e.g. GCP.PASHA O-sialoglycoprotein endopeptidase
(325 aa), fasta scores: opt: 626 z-score: 743.4 E(): 0,
36.2% identity in 312 aa overlap. 51.9% identity to
HP1584. Contains PS01016 Glycoprotease family signature,
and Pfam match to entry PF00814 Glycoprotease"
/codon_start=1
/transl_table=1
/product="putative glycoprotease"
/protein_id="CAB73771.1"
/db_xref="GI:6968778"
/translation="MKNLILAIESSCDSSAIIIDKNTLECKFHAKISOELHSITYGC
VPELARLHSEALPKMKOCKEHFKMLCAIVNEPGLSVILSGISMAKTLASAN
LPLIPINHLKGHIYLFLEKISLMDGIIIVSGHTEVILKDDASIELASTNDSE
GESFEDKAKMMNLGPGVGIIEMLAKAKIKNSFNPILKSKELASFGKNAVUL
EILKHEMLNEDTKAEIAYAPENTACDIIIMKLEIFNLYFKNGVGVGASANLNERS
RLONLCOKYANANKLAPLFCSDNALMIFARAADVAEYKKEVSVEEDILSPKKNKFSR
I"
misc_feature
complement(54255..55187)
/ gene="Cj1344c"
/ note="Pfam match to entry PF00814 Glycoprotease,
Glycoprotease family, score 423.70, E-value 1.7e-123"
misc_feature
complement(54849..54911)
/ gene="Cj1344c"
/ note="PS01016 Glycoprotease family signature"
complement(55190..56584)
/ gene="Cj1345c"
complement(55190..56584)
/ gene="Cj1345c"
/ note="Cj1345c, probable periplasmic protein, len: 464 aa;
41.2% identity to HP1075. Contains probable N-terminal
signal sequence"
/codon_start=1
/transl_table=1
/product="putative periplasmic protein"
/protein_id="CAB73772.1"
/db_xref="GI:6968779"
/translation="MKFELILEFITSIFALELDFSVGNGKSLDNNNYILFGGTOG
DEPGGFHASILSDYNTIKKTIIVAPNLFDSIIRSRGNNCDLNRKASISPKPD
YKTVQRIKELLILPEYSMTYINLHDGKGYTKPYTIDAMQPKRWGNSVIDTSEINAK
YFDLENIAOTVANSVNSLADPKHAHVLHNTQOELGDTQMELKALYFVYSNKAIFA
NEASKMLPVNLRAYYHLAIEYVLKTAGIEPSRDELTPQGVADKAIKKELEKLVFDDR

```

```

gene
complement(56586..57656)
CDS
/ gene="dxr"
complement(56586..57656)
/ gene="dxr"
/ EC_number="1.1.1.-"
/ note="Cj1346c, dxr, probable 1-deoxy-D-xylulose
5-phosphate reductoisomerase, len: 356 aa; similar to e.g.
dxr.BACSU 1-deoxy-D-xylulose 5-phosphate reductoisomerase
(EC 1.1.1.1) (388 aa), fasta scores: opt: 610 z-score:
706.2 E(): 4.3e-32, 40.4% identity in 349 aa overlap.
46.2% identity to HP0216"
/codon_start=1
/transl_table=1
/product="putative 1-deoxy-D-xylulose 5-phosphate
reductoisomerase"
/protein_id="CAB73773.1"
/db_xref="GI:6968780"
/translation="MILFGSTGIGVNALKTLAKNIPISALACGNALINEOIAIP
KPEVSIKDSKNKHLVHDVRFIGEGEJOILTECOOKLLNALYFAGLSTYKAKC
LGNALAIANKRESLYAGSEFLGKAFLEPDSHAAKLELBEKKNIAKUTIASGAFY
RYKIKDLNAVYQDAKLRHPNMNMAKITIDATANKLFEITIEAVHLYDFEIDALE
PRLSHVACERFNGASTVAFSRKADMKLAISAIIEKQDPTLEAVDSEKMPALFHP1
STRKVPYIFKLNTEFLKEPNLCVITIANEVGVNLEKSGFLDIACIFALDHFGV
PKISIEVEFEYDRTREYLSR"
gene
complement(57653..58378)
CDS
/ gene="cdsa"
complement(57653..58378)
/ gene="cdsa"
/ EC_number="2.7.7.41"
/ note="Cj1347c, cdsa, probable phosphatidate
cytidyllyltransferase, len: 241 aa; similar to e.g.
CDSA.ECOLI phosphatidate cytidyllyltransferase (EC
2.7.7.41) (249 aa), fasta scores: opt: 381 z-score: 465.5
E(): 1.2e-18, 40.7% identity in 182 aa overlap. 39.3%
identity to HP0215. Contains PS01315 Phosphatidate
cytidyllyltransferase signature, and Pfam match to entry
PF01148 cytidyllyltrans. Phosphatidate
cytidyllyltransferase"
/codon_start=1
/transl_table=1
/product="phosphatidate cytidyllyltransferase"
/protein_id="CAB73774.1"
/db_xref="GI:6968781"
/translation="MNTTPTISALVMGAIITIIIALIDQFINFIVPAVLLYSFSEA
KILPALENSITIPLAIFIGLSHKKALLGCIIVLLVGLVYKRSLSKLPALIYIP
SLPLALMOYLLDGMFLMWLITIVACOSGAIFGLKMGKTPSPSTSPNKTLEGVY
GGLICASVIGTILIGVFYSFMLSILCSFVAIFAVIGDLLESYFKRBAGVKDSODLIP
GHGVLEDRIDAVIIAVFMALL"
misc_feature
complement(57656..58375)
/ gene="cdsa"
/ note="Pfam match to entry PF01148 Cytidyllyltrans,
Phosphatidate cytidyllyltransferase, score 72.20, E-value
1.1e-17"
misc_feature
complement(57695..57775)
/ gene="cdsa"
/ note="PS01315 Phosphatidate cytidyllyltransferase
signature"
complement(58387..58725)
/ gene="Cj1348c"
complement(58387..58725)
/ gene="Cj1348c"
/ note="Cj1348c, probable coiled-coil protein, len: 112 aa;
lys-, glu-rich, highly hydrophilic, contains probable
coiled-coil from approx aa 50-80. No Hp match"
/codon_start=1
/transl_table=1
/product="putative coiled-coil protein"
/protein_id="CAB73775.1"
/db_xref="GI:6968782"

```

(21.1% identity in 563 aa overlap), and Cj0720c flac (25.6% identity in 203 aa overlap). Contains flm match to entry PF00669 Flagellin_N, Bacterial flagellin N-terminus, and Pfam match to entry PF00700 Flagellin_C, Bacterial flagellin C-terminus. The flaa and flab genes have a GC content of 30.5%.

/codon_start=1
/transl_table=1
/product="flagellin"
/protein_id="CA873765.1"
/db_xref="GI:6968772"
/translation="MGFRINTGALNAHNSVYVNAEQLDLSRLSSGLRINSAAD
ASGMATLDSRSQATLGAQINNGNDALIGILTADKAMDEQLTIDTIRKATQAAD
GOSLKTMTLQADINRLMEELDNIANTSPNGKOLLSGNFINOEFOIGASNOIRKAT
IGATOSKIGLITFRETNGRISGEVOFTLKNYNGIDPOFOKIVISTVSTGSGALVA
EINIKSADKTVARATFTVETRGIAVRAAGTTSDFAINGTIGVATIEDDNGALVA
AINSVKDTGVEASIDANGQLLTSREGRIKIDGNIGGAFINADMKENYGRISLVK
NDKRDILISGNSLSAGGATQFISOASVSLRESKREFNADAMFGSANGVYVAG
GYSVSAYMSAGSGSGSYSGSKNYSFGPANAIAISASQSLSTVYVNSGSGF
SSGTSLSQFATMTTAFGVDETAQVTLKGAMAVMDIAETATNLDQIRADIGSVON
QLQVTININIVTVQVNYKAESIRVDFAESANFSKNYINLAQSGSYAMSOANAQVON
VKILQ"

misc_feature
complement(45172..45387)
/gene="flab"
/note="Pfam match to entry PF00700 Flagellin_C, Bacterial flagellin C-terminus, score 128.80, E-value 1e-34"
complement(46312..46737)
/gene="flab"
/note="Pfam match to entry PF00669 Flagellin_N, Bacterial flagellin N-terminus, score 237.40, E-value 2e-67"
complement(46967..48685)
/gene="flaa"
complement(46967..48685)
/gene="flaa"
/note="Cj1339c, flaa, flagellin A, len: 572 aa; 80.7% identity to FLAA_CAMJF flagellin A (575 aa), fasta scores: opt: 2816 z-score: 2717.6 E(): 0, 80.7% identity in 575 aa overlap, 48.3% identity to HP0601. Also similar to Cj1338c (flaa), Cj0720c flac (24.5% identity in 220 aa overlap), and Cj0887c flab (20.2% identity in 560 aa overlap). Contains flm match to entry PF00669 Flagellin_N, Bacterial flagellin N-terminus, and Pfam match to entry PF00700 Flagellin_C, Bacterial flagellin C-terminus. The flaa and flab genes have a GC content of 37% compared to the overall genome GC content of 30.5%"

/codon_start=1
/transl_table=1
/product="flagellin"
/protein_id="CA873766.1"
/db_xref="GI:6968773"
/translation="MGFRINTVNAALNAKANADLNKSLDASLSRLSSGLRINSAAD
ASGMATLDSRSQATLGAQINNGNDALIGILTADKAMDEQLTIDTIRKATQAAD
GOSLKTMTLQADINRLMEELDNIANTSPNGKOLLSGNFINOEFOIGASNOIRKAT
IGATOSKIGLITFRETNGRISGEVOFTLKNYNGIDPOFOKIVISTVSTGSGALVA
EINIKSADKTVARATFTVETRGIAVRAAGTTSDFAINGTIGVATIEDDNGALVA
AINSVKDTGVEASIDANGQLLTSREGRIKIDGNIGGAFINADMKENYGRISLVK
NDKRDILISGNSLSAGGATQFISOASVSLRESKREFNADAMFGSANGVYVAG
GYSVSAYMSAGSGSGSYSGSKNYSFGPANAIAISASQSLSTVYVNSGSGF
SSGTSLSQFATMTTAFGVDETAQVTLKGAMAVMDIAETATNLDQIRADIGSVON
QLQVTININIVTVQVNYKAESIRVDFAESANFSKNYINLAQSGSYAMSOANAQVON
VKILQ"

misc_feature
complement(47036..47251)
/gene="flaa"
/note="Pfam match to entry PF00700 Flagellin_C, Bacterial flagellin C-terminus, score 130.80, E-value 2.5e-35"
complement(48176..48601)
/gene="flaa"
/note="Pfam match to entry PF00669 Flagellin_N, Bacterial flagellin N-terminus, score 237.50, E-value 1.9e-67"
complement(48782..50599)
/gene="Cj1340c"
complement(48782..50599)

CDS

/gene="Cj1340c"
/note="Cj1340c, unknown, len: 605 aa; no Hp ortholog, but 28.5% identity to HP0114. A member of the 1318 family of proteins. Similar to Cj1341c (83.4% identity in 598 aa overlap), Cj1337 (32.5% identity in 625 aa overlap), Cj1333 (30.4% identity in 615 aa overlap), Cj1336 (32.0% identity in 579 aa overlap), Cj1318 (31.4% identity in 641 aa overlap), and Cj1334 (34.1% identity in 618 aa overlap)"

/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1340c (1318 family)"
/protein_id="CA873767.1"
/db_xref="GI:6968774"
/translation="MNLEKNIOALLSGVNEPLGNKLNFTQNTCSRPNIDENLNF
DKTHNVMYENLEEFENFPOSILEKTPRYPCITGIGNALLIKNAKHYKHLFVE
SEIEPLFLASTIDLSIELKAYKVYIPDCYAKDEIDIMAFDOQSILEVLSLEMTI
SSHLYLVETSLISLNELCIKSASVAIRNADITCFILPLTHGQFLQNTFSMESTPE
QRIISERKNFEENAVIVSAGPSLAKQDPLKAYODKAVIFCADALSMLEKEGITIPY
VTNLDFTDLAMKFPONKENKOSIIALECATHPVIRASLNCMIVLRKALYQREN
LNDGVIDTGGVSHFSYTLALAGFNIIIMIGODLAFDEGNSHSGPFGEGFSGE
ENIDKIKVTYAGKEVLTHTNMDYRIKIEYFACMEOKAKFPNAREGARRNFEPE
LSRECECEKLTREKPOPELPSITKRSKLYKREKIOKODNAKRLDIALAK
QIENILSKDFILPLELEKYVQNTENFNSLDEDEFTIDETLNGAFVARGMTADVL
KLHIDKTHFTITAYIKAYHMLVFIETLEKOKYKSLSKV"
complement(50596..52419)
/gene="Cj1341c"
complement(50596..52419)
/gene="Cj1341c"
/note="Cj1341c, unknown, len: 607 aa; no Hp ortholog, but 28.4% identity to HP0114. A member of the 1318 family of proteins. Similar to Cj1340c (83.4% identity in 598 aa overlap), Cj1334 (34.6% identity in 622 aa overlap), Cj1336 (30.8% identity in 578 aa overlap), Cj1333 (31.2% identity in 616 aa overlap), Cj1318 (30.2% identity in 640 aa overlap), and Cj1337 (32.0% identity in 628 aa overlap)"

/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1341c (1318 family)"
/protein_id="CA873768.1"
/db_xref="GI:6968775"
/translation="MTLEKNIOALLSGVNEPLGNKLNFTQNTCSRPNIDENLNF
DKTHNVMYENLEEFENFPOSILEKTPRYPCITGIGNALLIKNAKHYKHLFVE
SEIEPLFLASTIDLSIELKAYKVYIPDCYAKDEIDIMAFDOQSILEVLSLEMTI
SSHLYLVETSLISLNELCIKSASVAIRNADITCFILPLTHGQFLQNTFSMESTPE
QRIISERKNFEENAVIVSAGPSLAKQDPLKAYODKAVIFCADALSMLEKEGITIPY
VTNLDFTDLAMKFPONKENKOSIIALECATHPVIRASLNCMIVLRKALYQREN
LNDGVIDTGGVSHFSYTLALAGFNIIIMIGODLAFDEGNSHSGPFGEGFSGEENI
DKIKVAYAGKEVLTHTNMDYRIKIEYFACMEOKAKFPNAREGARRNFEPELS
RECECEKLTREKPOPELPSITKRSKLYKREKIOKODNAKRLDIALAKQIENILSKDFILPLELEKYVQNTENFNSLDEDEFTIDETLNGAFVARGMTADVL
KLHIDKTHFTITAYIKAYHMLVFIETLEKOKYKSLSKV"
complement(52429..53670)
/gene="Cj1342c"
complement(52429..53670)
/gene="Cj1342c"
/note="Cj1342c, unknown, len: 413 aa; no Hp match. A member of the 617 family of C-j. proteins containing homopolymeric tracts; similar to Cj1310c (37.3% identity in 424 aa overlap), Cj1305c (38.5% identity in 423 aa overlap), Cj0617/Cj0618 (49.7% identity in 191 aa overlap), Cj0617/Cj0618 (49.7% identity in 191 aa overlap), Cj1306c (38.3% identity in 426 aa overlap). Contains C(9-10) polymorphic region at aa 188. C(9) consensus gives this ORF, C(10) gives truncation after a further 19 aa"

/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1342c (617 family)"
/protein_id="CA873769.1"
/db_xref="GI:6968776"
/translation="MIVSKAYEIDSCDDVDELGTRRSKLEFKLYADYDEKTPALFVY
OGMGDDISSIOYGVDMVROFNFAVGVNHCIGNRSPSSFFMDIDKLILKASCE

CDS

```

overlap), Cj1337 (48.6% identity in 636 aa overlap),
Cj1334 (41.8% identity in 644 aa overlap), Cj1340c (30.48
identity in 615 aa overlap), Cj1341c (31.2% identity in
616 aa overlap) and Cj1335/Cj1336 (94.2% identity in 52 aa
overlap/71.7% identity in 597 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1333 (1318 family)"
/protein_id="CAB73760.1"
/db_xref="GI:6968767"
/translation="MGIMMFTPTQKELFNNKINIALSNLFLKESIKKESKFEIILG
KDNILNKKSDNMTGLYEMVIDELSNLNTYNDKYLILPVLYFGNGNGLFALLO
KNHOHIVLEPKDIEITWVFIHILDSNEELQSAARMILITQTSIDILFSPNCSKRPF
QSRHIVLEKMSHYERHEVLELNKKLVDPFSDIISHGNDPUDALQIGIEFVYNL
POMITHSYKELSKRGISDTALIVSGPSLTQPLPKKVASKATIFCADSSYPIL
AKHNKIDYVLSLRIPLTSEFNNDGEPKDLIFLKSIVHPHTTKYLQKNRNM
LVSTYASFIVYLKDDGYEFNMGSVANMFLIALHLKKNIVLIGDLAAKQGLSH
TKDYSNLDKHEGDRKNTYTOAGNGKVESFPWTLFRHNEODVANAANKNYIT
TYNCTGGARIEGTIEKPFIMACENILHKDLNKPFEKILPELUNKONEILKAYK
YOSTKHCRPSNKRITKSYDKIKNSFMSLQNSQENETLKEITKIDIKIDITIDELVNT
OKDLMLIGPLTQPELNLARIYVNLNPKTKEDAKNSILWIKHELMELVYGHKAKO
ENALIKNIPLEELKERRKLDKMERVR"
39370..41229
/gene="Cj1334"
39370..41229
/gene="Cj1334"
/note="Cj1334, unknown, len: 619 aa; no Hp ortholog, but
34.8% identity to Hp014. A member of the 1318 family of
proteins. Similar to Cj1333 (41.9% identity in 645 aa
overlap), Cj1318 (41.0% identity in 647 aa overlap),
Cj1336 (41.9% identity in 590 aa overlap), Cj1337 (43.8%
identity in 628 aa overlap), Cj1341c (34.3% identity in
621 aa overlap), and Cj1340c (34.0% identity in 618 aa
overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical proteoetin Cj1334 (1318 family)"
/protein_id="CAB73761.1"
/db_xref="GI:6968768"
/translation="MRELEFLKNTQALFEVDEFLACTLSIKYLTFTALIDENGINKR
KIDIFLEPNKELLENLTEKTEYKNYPVLFYFGNGMFEYKTLCKNKHIIIE
DNLEILLAHLFDFSEELKEQLILFTYPTNINATLTFTYIIOKSVKINFLFIH
NDFYQGYSTQIOMINQIEMIRFVILNKGNDPDSLVGIKHTLDLPRMLNNGTIO
EPLKERAKVENATIVSTGPSIKQPLKKYANKATIFCADSSYPILAKHGKIPDY
CMLEKRDIVSKCEDNDGPNKGIKFLFVAVHEVDFLEKDRAMLVLRPNTFA
SLKIDETGIVGHSVSNMILEGALRFENILFTIGDLAGEDGSSHPREHNGSOG
EELRGEKVTYLAGGKRVQTLTMNLFROAFEDIFWAKERLNTTYNCTGGARIEG
TIEKPFIMACENILDKNKPFPKPKLAKERLEKTKRYLOKSILSKERTIKT
OTOLKRLRYLEKNDKMFOTLEKIKNDLNLKFEKKTLFNLCOAIFYHNEEILK
FEVLNTRKOKENILDFLKIQHNFIQGLGDLTONKTIKESLEMNPDITIK"
41234..43191
/note="repeat 3; identical to 1246669..1248628 (approx.
positions)"
41242..41424
/gene="Cj1335"
41242..41424
/gene="Cj1335"
/note="Cj1335, unknown, len: 60 aa; no Hp match. A member
of the 1318 family of proteins. Identical to the
N-terminus of Cj1318 (100.0% identity in 59 aa overlap)
and similar to Cj1333 (94.2% identity in 52 aa overlap).
Contains a G(9-10) polymorphic region at aa 57..G(9)
(consensus) gives this ORF, G(10) would extend the ORF by
18 aa. G(8) or G(11) (not seen in shotgun) would allow
translation into the downstream ORF Cj1336, giving a
protein similar to other members of the 1318 family"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1335 (1318 family)"
/protein_id="CAB73762.1"
/db_xref="GI:6968769"
/translation="MTFTPTQKELFNNKINIALNNILLKESLKEIKSKFELLIGKNL
DINKDTSIKNNGGL"

```

```

gene
41408..43189
/gene="Cj1336"
41408..43189
/gene="Cj1336"
/note="Cj1336, unknown, len: 593 aa; no Hp ortholog, but
34.0% identity to Hp014. A member of the 1318 family of
proteins. Identical to Cj1318 (except the C-term, 100.0%
identity in 592 aa overlap), and similar to Cj1333 (71.7%
identity in 597 aa overlap), Cj1334 (42.3% identity in 591
aa overlap), Cj1337 (48.6% identity in 586 aa overlap),
Cj1341c (30.8% identity in 578 aa overlap) and Cj1340c
(32.0% identity in 579 aa overlap). Contains a G(9-10)
polymorphic region at aa 2..G(9) (consensus) gives this
ORF, G(10) would allow translation to start at aa 21. G(8)
or G(11) (not seen in shotgun) would allow translation
from the upstream ORF Cj1335, giving a protein similar to
other members of the 1318 family"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1336 (1318 family)"
/protein_id="CAB73763.1"
/db_xref="GI:6968770"
/translation="MGGYNNELLYQDPIKELQTMNTYNDKYLILPVLYFGNGCI
LEKALLNKNHOHIVPEKDIETITWVFIHILDSNEELQSAARMILITQTSIDILF
CSKRPFOFSRIVLELMSHYERHEVLELNKKLVDPFSDIISHGNDPUDALQIGI
EQFVYNLPOMITHSYKELSKRGISDTALIVSGPSLTQPLPKKVASKATIFCA
DSSYPILAKHGKIPDYCMLEKRDIVSKCEDNDGPNKGIKFLFVAVHEVDFLEK
GRNVLITOKYLAAPYIINLKDSYAAGSLVAHTLSVATYLSKRTIFICODLAVAE
NGNSHPDYOXSAVESQMYRHITETAYGNGKVEYTHSLFKNWENDEPIATKRM
GITTYNCTGGARIEGTIEKPFIMACENILDKNKPFEKILPELUNKONEILKAY
KVSOSIHCRPESKILSNDFENNIONITYLNKKNENDLNLARIKIDERNKLENTKOQ
DLYEILQPLRTQPELNLARIYVNLNPKTKEDAFNKSIIIMKHELMELVYGHKAKO
ALIKNIPLEELKERRKLDKMERVR"
41410..41418
/misc_difference
41410..41418
/gene="Cj1335"
/note="polymorphism G(9-10)"
43186..45072
/gene="Cj1337"
43186..45072
/gene="Cj1337"
/note="Cj1337, unknown, len: 628 aa; 35.0% identity to
Hp014. A member of the 1318 family of proteins. Similar to
Cj1333 (48.6% identity in 636 aa overlap), Cj1318 (48.1%
identity in 643 aa overlap), Cj1336 (48.6% identity in 586
aa overlap), Cj1334 (43.6% identity in 628 aa overlap),
Cj1340c (32.2% identity in 624 aa overlap), Cj1341c (31.9%
identity in 627 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1337"
/protein_id="CAB73764.1"
/db_xref="GI:6968771"
/translation="MRENNFNKRLKALSGEVNLLKRLKEDLKELEREFPTGKNL
DINIKRRNLKMYQDPIKELKELFEPKDFTRPVLFYFGNGILYKTLIDQALIK
RIITFEKELITFALNFDPSKDSLSGRLLIHHDDINLPKDKVRLIGDLEFRSY
NLHIANDFEYHKEDILKLNKNQIIRKNHLMGNDPKDALQIGIEFVYNLPSMITH
PSYKELSKRGISDTALIVSTGPSLTQPLPKKVASKATIFCADSSYPILAKHGK
PDYVCMLEKRDIVSKCEDNDGPNKGIKFLFVAVHEVDFLEKGRNVLITOKYLA
PVARGLIDYGYINAGMSVSHVAEYELAENIGHNNIITLIDODLAAKDCGTHSQEPH
ANLHNGYENDUDRFTTAYGNGKVOSSSETVLPFOITFENFLAFSKTYNCTEGA
RISALIEKPRELCELDLKNKKRKKRKYLVNTRKEQVKGILAYOKIKNNMISLNF
KKECKYOKOIHNTGKRNLSLEQIONIDIKIKELSNKKRYFLQELIGLPTLHNEG
ILPLVLYKDIKDESDKONKLFAMWYAHLEALIENTIELLEVDQRKRIALPLDIFEK
KRAL"
complement(45103..46821)
/gene="flaB"
/gene="flaB"
complement(45103..46821)
/gene="flaB"
/note="Cj1338c, flaB, flagellin B, len: 572 aa; 80.9%
identity to FLAB_CAMJ6 flagellin B (575 aa), fasta scores:
opt: 2812 z-score: 2717.5 E(-): 0.80 98 identity in 575 aa
overlap. 52.1% identity to Hp015. Also similar to Cj1339c
flaB (93.9% identity in 572 aa overlap), Cj0887c flabB

```

```

/product="putative N-acetylglucosamine-6-phosphate
2-epimerase/N-acetylglucosamine-6-phosphatase"
/protein_id="CAB73755.1"
/db_xref="GI:6968762"
/translation="MSKRKICIVSATRAEWLRLNLCHLEIONDKDLSLQIITGAHIS
PERGLYIKETKEFKTKIKIPILLANDKISLCSMSLAFSAFSDAFEDKPDWVIL
GDRYEMLSVASVCLMHIPLVHLCGGELTGADDLSIRHSISKMSHLFPWSHELYKRR
LLOQEEERKSVASGLSLASTIINKMNFLEKLEALEMKLDELILTYHPLTANK
NTOKEITLKLKDLTKLASLIETKANADENGLIINELIONYCOKNSHAKALPDNLGS
OKYLSLTKIAKAMIGNSSGISEPFKTCINIGDRKGKGRITRONITDSEINDLDA
FEKLESKEPKNLKNFPNDKNGKNPKIKITCKLVNLDITLHKNFIDL"
34027..34182
/gene="Cj1329"
/feature="Pfam match to entry PF00571 CBS, CBS domain, score
13.10, E-value 0.72"
34027..35052
/gene="Cj1329"
34027..35052
/gene="Cj1329"
/feature="Cj1329, probable sugar-phosphate nucleotide
transferase, len: 341 aa; similar in C-term (120 aa - end)
to many predicted transferases, and to MGL_YEAS7
mannose-1-phosphate guanylyltransferase (361 aa), fasta
scores: opt: 304 z-score: 343.2 E(): 8.1e-12, 28.9%
identity in 225 aa overlap. N-term has some similarity to
CBS domains in nucleotide dehydrogenases e.g. IMDH_BACSV
inosine-5'-monophosphate dehydrogenase (513 aa), blastp
scores: E=0.00034, 30% identity in 116 overlap. N-term
contains 2x Pfam match to entry PF00571 CBS, CBS domain,
C-term contains Pfam match to entry PF00483
NTP-transferase, Nucleotidyl transferase. No Hp match.
Also similar in C-terminus to Cj1423c (35.6% identity in
216 aa overlap)"
/codon_start=1
/translation="1
/feature="putative sugar-phosphate nucleotide transferase"
/protein_id="CAB73756.1"
/db_xref="GI:6968763"
/translation="MDINKLTPDSSIKKELKIVGQERVRLGIYVKKRPLGVISD
SNIRKALISGKTLKSDIKDIYTKNPITIKENTSEKDLIKTSKTDIYDPVLDKCOI
LSIKSISLILKANDNSIIMAGLGSMLKELTDTPRPMKLGKRPILSEIIOBLKQO
NENFLEPCVAKKQIILEDYFQKQKQEVKISYIKERKLTAGALSILKQEFKESFY
MNADITLEDNALNLAHRSKALMSVCVEFEQIPIGVITQKGFIEIEKPTOK
FLVSAGIYVLENIENLILAKNEYLDMPELILKYLQKGVNTYIINDYIDIGRDEPL
KANEDFK"
34213..34374
/gene="Cj1329"
/feature="Pfam match to entry PF00571 CBS, CBS domain, score
21.80, E-value 0.016"
34384..35043
/gene="Cj1329"
/feature="Pfam match to entry PF00483 NTP-transferase,
Nucleotidyl transferase, score 112.60, E-value 7.8e-30"
35049..35954
/gene="Cj1330"
35049..35954
/gene="Cj1330"
/feature="Cj1330, unknown, len: 301 aa; some similarity to
predicted oxidoreductases e.g. TR:O26961 (EMBL:AE00063)
Methanobacterium thermoautotrophicum MRH75 (predicted
3-chlorobenzoate-3,4-dioxygenase dhydrogenase related
protein) (318 aa), fasta scores: opt: 199 z-score: 244.0
E(): 2.7e-06, 24.7% identity in 267 aa overlap, and
TR:Q53739 (EMBL:X92429) Streptomyces alboniger predicted
oxido-reductase (338 aa), fasta scores: opt: 135 z-score:
167.8 E(): 0.048, 27.0% identity in 196 aa overlap. No Hp
match"
/codon_start=1
/translation="1
/feature="hypothetical protein Cj1330"
/protein_id="CAB73757.1"
/db_xref="GI:6968764"
/translation="MKVLIFGSGICKHPLALKNLHNEVSLSLSAKKEFKNTPYI

```

```

RSLEKCDLNEFDLFIANTTHTHTLTKLDELVKDKTKLIVKPEKSONFTSSKNH
IYVAYLQRFNPVIALKRLKGEKITYFASVYCNLSYHPMALDQRYVSAKKEGCV
LIDLSHEDLAFPLFKDLIELYSQNAKISELITSDSDFALAKSODAKIHLDFE
AKENKREIITITLESKRALINKIKELIYHKNSOKILNEENDITKILQNLHHAVERK
DKELCDLKQALKVLELDEVRKNG"
35947..36654
/gene="p1mb"
35947..36654
/gene="p1mb"
/feature="p1mb"
/EC_number="2.7.7.43"
/feature="Cj1331, p1mb, neuA3, probable acylneuraminate
cytidyllyltransferase (CMP-N-acetylneuraminate acid
synthetase), len: 235 aa; 83.8% identical to P1MB_CAMCO
posttranslational flagellin modification protein B (235
aa), and similar to e.g. NEUA_NEIME acylneuraminate
cytidyllyltransferase (EC 2.7.7.43) (228 aa), fasta scores:
opt: 357 z-score: 458.8 E(): 3.8e-17, 30.1% identity in
226 aa overlap. No Hp ortholog. Also similar to e.g.
Cj1311 neuA2 (32.9% identity in 234 aa overlap) and Cj1143
neuA (34.7% identity in 219 aa overlap)"
/codon_start=1
/translation="1
/feature="acylneuraminate cytidyllyltransferase (flagellin
modification)"
/protein_id="CAB73758.1"
/db_xref="GI:6968765"
/translation="MARILTCARGSGSKGVKNKNIKMLKLEMAVSIIOANSKLF
KHIVISDSEIASVAKYAEVYFKREAHLANDRTPKLPYMRALLRSEHRTCE
LTLIDLSAPLRSLDKRAKESFEVDNSNLITVAPKRNPFNLVLEIKNKVKSK
EGNFTTRQSAFKCYDMNASIYIFKRDYLLENDSVFKNGLTFWDESTADIDSELF
KIVEFLISLNLSPKDE"
36654..37424
/gene="p1ma"
36654..37424
/gene="p1ma"
/feature="p1ma"
/feature="p1ma, ptma, probable oxidoreductase, len: 236 aa;
79.3% identical to PTMA_CAMCO posttranslational flagellin
modification protein A (256 aa), and similar to e.g.
DHG2_BACME glucose 1-dehydrogenase II (261 aa), fasta
scores: opt: 236 z-score: 281.8 E(): 2.1e-08, 24.1%
identity in 261 aa overlap. No Hp ortholog. Contains fam
match to entry PF00106 adh_short, Alcohol/other
dehydrogenases, short chain type, and Pfam match to entry
PF00678 adh_short_C2, Short chain dehydrogenase/reductase
C-terminus"
/codon_start=1
/translation="1
/feature="putative oxidoreductase (flagellin
modification)"
/protein_id="CAB73759.1"
/db_xref="GI:6968766"
/translation="MLENKIIFVAGACGRIGKALCEKILSKGPIIADINKECLKL
QEWLETFKTKLISLELDITKQESLQALQKSDERKIDAFVNSSPFEKDMGKTPY
VELKEDQICSLNLHAGFMIAQOEKPEKQGYGNIMLSSIMGYAPKFNENYGT
SMQSLSEYSVIKAGINRNHMSMLAKELFNOMIRNTLASGILDNQNELFKAVRKCA
SKGMLADADICGTVFLSDSESKFTIGQTLVDDGMCL"
36666..37265
/gene="p1ma"
/feature="p1ma"
/feature="p1ma match to entry PF00106 adh_short,
Alcohol/other dehydrogenases, short chain type, score
65.00, E-value 1.6e-15"
37311..37403
/gene="p1ma"
/feature="p1ma match to entry PF00678 adh_short_C2, Short
chain dehydrogenase/reductase C-terminus, score 35.10,
E-value 1.6e-06"
37412..39373
/gene="Cj1333"
37412..39373
/gene="Cj1333"
/feature="Cj1333, unknown, len: 653 aa; no Hp ortholog, but
34.6% identity to HP0114. A member of the 1318 family of
proteins. Similar to Cj1318 (72.5% identity in 658 aa

```



```
misc_feature
/translation="MKNVIIGAGGFTRELYSYLDANEIIGYIDIOENPFDKYL
GNEIDPFDKLIDKASFPALGVOGQINLRKILVLSKSSCFITFIHPQSVSEAKIQ
GVIVCPPTVINANSNIGDFVLCNIYSIAHDCVGEISLSPYATLNGSSIGKNCFL
ATRVSLPCVNIEDNCIVSR"
28866..28943
/gene="Cj1321"
/feature="Pfam match to entry PF00132 hexapep, Bacterial
transferase hexapeptide (four repeats), score 14.60,
E-value 0.027"
28974..29078
/gene="Cj1321"
/feature="Pfam match to entry PF00132 hexapep, Bacterial
transferase hexapeptide (four repeats), score 0.90,
E-value 2e+02"
29202..29522
/gene="Cj1322"
29202..29522
/gene="Cj1322"
/feature="Cj1322, unknown, len: 106 aa; no Hp match"
/codon_start=1
/translation="hypothetical protein Cj1322"
/protein_id="CAB73749.1"
/db_xref="GI:6968756"
/translation="MKEINIIDFGMLKQISALFYLGIEGVYKSKLNIYEERKQI
KLQKRIKIDFSNFKGINIYQHIEDLKNSLTLESINEDLNLRKELMQILRDNNIVSR
VMI"
29510..29896
/gene="Cj1323"
29510..29896
/gene="Cj1323"
/feature="Cj1323, unknown, len: 128 aa; no Hp match"
/codon_start=1
/translation="hypothetical protein Cj1323"
/protein_id="CAB73750.1"
/db_xref="GI:6968757"
/translation="MDLNCDFIHFENITKILIELCGSNLERFPLDLKLGPHII
CSKNGRALANLILNEISSEFKIIEKYDYNKRCOTYADLYEKRNLIIDITGIE
LCNKICKMLKEDDENFHPKIFKKAL"
30013..31134
/gene="Cj1324"
30013..31134
/gene="Cj1324"
/feature="Cj1324, unknown, len: 373 aa; some similarity to
TR:P72140 (EMBL:U50396) Pseudomonas aeruginosa MBPG
(11popolysaccharide gene cluster) (376 aa), fasta scores:
opt: 166 z-score: 199.7 E(): 0.0008, 20.1% identity in 378
aa overlap. No Hp match. Also similar to Cj1316c"
/codon_start=1
/translation="hypothetical protein Cj1324"
/protein_id="CAB73751.1"
/db_xref="GI:6968758"
/translation="MICYDHCVMPTPGINFTKDESKNICSACINRKNENIDYKA
RPELEVLCDKRYRRNNGKFEYDCATAVSGKDSHVOVHIMEKLMNPILFSVEENFT
MTEAGKNNLNLSEFEGCHIIISLPDITQKVMKTFEKGKPTWFTDLIYSFPA
MALKNTPLLVYGENVSYEGSDTEEPKSAKEIPLNGVADSLNINEFTIDEIKENL
OLFENPKDKLDKLPILYISYFVKWNSYNTIASKRGFTIDEGEDRMTCAENEDOV
DSIGYILHAMKYPKEGHACADVAARVRYGILSRKAIELVQKRDHKLDNKCVEDF
CNFGISKTFTWKVIEKHYNMDFYKNDPGEKRLKNLQ"
31152..31421
/gene="Cj1325"
31152..31421
/gene="Cj1325"
/feature="Cj1325, unknown, len: 89 aa; no Hp match. Contains
polyporphic sequence G(9-10) at C-terminus; G(10)
(consensus) gives this translation stop, G(9) allows
translation into the downstream CDS Cj1326. Similar to
Cj0170 (73.8% identity in 61 aa overlap), which also
contains a polymorphic sequence allowing possible
translation into Cj0171"
/codon_start=1

misc_difference
31404..31413
/gene="Cj1325"
/feature="polymorphism G(9-10)"
31405..31827
/gene="Cj1326"
31405..31827
/gene="Cj1326"
/feature="Cj1326, unknown, len: 140 aa; no Hp match. Contains
polyporphic sequence G(9-10) at N-terminus; G(9) allows
translation from the upstream CDS Cj1325, G(10)
(consensus) would only allow this CDS to start at aa 22.
Similar to Cj0171 (51.9% identity in 135 aa overlap),
which also contains a polymorphic sequence allowing
possible translation from Cj0170"
/codon_start=1
/translation="hypothetical protein Cj1326"
/protein_id="CAB73753.1"
/db_xref="GI:6968760"
/translation="GGGCKIIEPNSSIKGLFDENMDIIFANOSLYIPLKLNIL
EYELLNTGILFATMSKKNYFSPSHQKEENGLSKVINRNLNETSIHRIIDAED
LENLFQPERTLGLDIDPILNFTNFGSAHHYITGIKK"
31867..32871
/gene="neuB2"
31867..32871
/gene="neuB2"
/feature="neuB2"
/feature="4.1.3.-"
/feature="Cj1327, neuB2, N-acetylneuraminic acid synthetase,
len: 334 aa; similar to e.g. TR:Q45675 (EMBL:U05248)
E.coli polysialic acid gene cluster region 2 neuB (346
aa), fasta scores: opt: 1112 z-score: 1285.6 E(): 0.51.5%
identity in 336 aa overlap, and TR:Q57265 (EMBL:U40740)
Neisseria meningitidis N-acetylneuraminic acid synthetase
(BC 4.1.3.-) (349 aa), fasta scores: opt: 624 z-score:
724.6 E(): 4.6e-33, 34.5% identity in 333 aa overlap"
/codon_start=1
/translation="N-acetylneuraminic acid synthetase"
/protein_id="CAB73754.1"
/db_xref="GI:6968761"
/translation="MKTILINEAGVNHNGDLNLRKLEIADSGADFKFQSRKAK
NCISTRKARAPQLTANDESQLOMVQKLELDLAHRELILHAKCNIAFSTPDL
ESVDLINEGLKIFKIPSGEITNPLYLKAIKANKIILSTGANIAGEIEALNVLC
NGAKRONITLLCTEYEPAPFNEVNLKAMQSLKDAFKLDVGSDBHTRGHISLAVAL
GACVIEKHFTLDKNSGPDHAKSLPEQELKMCCTOIRIOKMGGCIKASKSPKNI
NIVRSIYAKKIKKIGELFSEGNLTTKRPANCISMRREEFLKATATNRYKEDLIRE"
32864..34018
/gene="neuC2"
32864..34018
/gene="neuC2"
/feature="neuC2"
/feature="5.1.3.-"
/feature="Cj1328, neuC2, probable
N-acetylglucosamine-6-phosphate
2-epimerase/N-acetylglucosamine-6-phosphatase, len: 371
aa; similar to e.g. TR:Q47400 (EMBL:M84026) E. coli neuC
(391 aa), fasta scores: opt: 779 z-score: 889.7 E(): 0.
36.1% identity in 388 aa overlap, and TR:Q57141
(EMBL:X78068) Neisseria meningitidis putative
N-acetylglucosamine-6-phosphate 2-epimerase (EC 5.1.3.-) /
N-acetylglucosamine-6-phosphatase (EC 3.1.3.-) (377 aa),
fasta scores: opt: 561 z-score: 642.7 E(): 1.7e-28, 33.5%
identity in 391 aa overlap. No Hp match. Also similar to
Cj1142 (34.6% identity in 384 aa overlap)"
/codon_start=1
/translation="N-acetylglucosamine-6-phosphate 2-epimerase"
```

```

IIMGESPOLEIYGSPASSKNNKILGRMLEPFGGLGNRASDMLGVNGITEKDLFLTYT
PSEDELRGVGTGLFGLGYFKMDYKILKISRYGFLTLDPHPEVTEYENPNDCCSN
HVHDLTKYKSGEGRATDNACDIRGLYISREKGVYLAQKDGKPPKKAIKKYLEHSG
FSSEEROKIVDSFTNKKIKRDNCKGFIIDYDGSIVKRDCVYK"
23385..24416
/gene="neub3"
/db_xref="GI:24416"
/EC_number="4.1.3.-"
/feature="Cj1317, neub3, N-acetylneuraminic acid synthetase,
len: 343 aa; 97.6% identical to TR:052907 (EMBL:AJ000855)
C. jejuni neub, and similar to e.g. TR:057265
(EMBL:U040740) Neisseria meningitidis N-acetylneuraminic
acid synthetase (EC 4.1.3.-) (349 aa), fasta scores: opt:
584 z-score: 664.1 E(): 1.1e-29, 35.1% identity in 350 aa
overlap, 49.5% identity to HP0178. Also similar to Cj1327
(36.3% identity in 336 aa overlap), and Cj1141 (34.7%
identity in 340 aa overlap)"
/codon_start=1
/transl_table=11
/product="N-acetylneuraminic acid synthetase"
/protein_id="CAB73744.1"
/db_xref="GI:6968751"
/translation="MOIGNFNTDKKVFITAEISAHNAGSLEMAKLSIKAKKAGADAI
KIQYTPSLTLNSDKEDFIKGLMDKRLKYLESAKPYEMHSQIFETAEQNEGIL
CFSSPPAKEDVEFLKRFPIAKIASFEANDENFVILAKKERTPTVSTGISTEEELF
KICEIFKEEKNDLVEKCTSTYPTAIEDMNLKGIYSLKFKFNVEGSLDSHGFIAP
VMAVALGARVIEKHMIDKSTESDSKSLSDPEFAWDAYDAVROAFSALGDCKLIDE
KYLKNRPARSLYASKDKIKGEMSEENKSVRPSGILHPKTYQELIGKRAKSKDKIKFG
DALQGDPO"
24405..26364
/feature="repeat 3; identical to 1263498..1265455 (approx.
positions)"
24413..26362
/gene="Cj1318"
24413..26362
/gene="Cj1318"
/feature="Cj1318, unknown, len: 649 aa; no Hp ortholog, but
33.9% identity to Cj1335/Cj1336 (100.0% identity in 59
aa overlap/100.0% identity in 592 aa overlap), Cj1333
(72.5% identity in 658 aa overlap), Cj1334 (41.5% identity
in 648 aa overlap), Cj1337 (48.1% identity in 643 aa
overlap), Cj1341c (30.2% identity in 640 aa overlap),
Cj1340c (31.4% identity in 641 aa overlap). Contains
G(10-11) polymorphic tract at aa 57, G(11) (consensus)
gives this ORF, G(10) causes truncation after 1 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1318 (1318 family)"
/protein_id="CAB73745.1"
/db_xref="GI:6968752"
/translation="MTPTPOKELFNKNTEALNNILKESLKEIKSKRELLGKDLN
DINLKDTSKNGGGYNDLIDYPIKELQMLTNYNDKYLIPVLYFGGNSILF
KALLQNNKHQHLVEKEDIEITWIFHLDFSNELQNNLAWLNDKLDQDQYELCS
SKPFQSNKIFYLELMSHYERFHEDEVLKLAENFNKILNNGNDPKDQDYLQIEO
EYVNLPMQTHPSYKELSKRNLSGPAIIVSTGSLFKOLPLKRYASAKTIFADS
SYPLIAKGKIPDYCMLEKRETELTAEPFNHDEGFDKDIYFCAGVHKAIEYKGR
NLVITOKVLAIPPYINLKDFAAGLSVAHRLSLATLTSIKRNIIFIGODLAVANG
NSHPDQNSANYESOMTEHILTEAVYGNGKVEHSIWLEKKNWENENIIPTRMGI
TYYNCTEGGARLEGTIERKFLWACENLMDKLNKPEKLEPISLKNQNEKIPKAYKY
YOSIKHCDKFSKLSNDFNNIIONIYLNNKKNENDLAIIRKIDEPKNNLETKQMDL
YEIOLPRLQFELNARIYVLPKTKEDAFNKSILMIKEHLEFMELVYGHKAQENAL
IKNILLPEKLERKLDKMERVRR"
24581..24591
/misc_difference
24581..24591
/gene="Cj1318"
/feature="polymorphism G(10-11)"
26359..27330
/gene="Cj1319"
26359..27330
/gene="Cj1319"
/feature="Cj1319, probable nucleotide sugar dehydratase, len:
323 aa; similar to e.g. TR:054116 (EMBL:L37354) thymidine
diphosphoglucose 4,6-dehydratase (329 aa), fasta scores:
opt: 526 z-score: 600.3 E(): 3.9e-26, 29.5% identity in
315 aa overlap. No Hp ortholog. Contains Pfam match to
entry PF00106 adh_short, Alcohol/other dehydrogenases,
short chain type"
/codon_start=1
/transl_table=11
/product="putative nucleotide sugar dehydratase"
/protein_id="CAB73746.1"
/db_xref="GI:6968753"
/translation="MKNIIVTGAQGTGSHLCESLVKKGFYKRAISQVSNFNGHLE
KSPFLKMDVEISGDLDSFCEKRTKNIDAIIFHLGALNIPYSATPQSYDPTNNGT
LMLLEAKNNEISHFHTSTSEYGTAFVPIIDKELQDPOPSYSKIAADMALSY
YNSFNILNARIPFVYNGPQSAKRAIIPITIIQISGAKEIKGIDLSPKRDLNFIYDT
GSGFISLNLVIRFGEVYNGISGVESQMOVLQIKIILSKYKTIIDDEGRLPKNSEV
PRLCCDANKTKATNQSNISLSLEGLRQSIKFKENLEKVELYNV"
26362..26895
/gene="Cj1319"
/feature="Pfam match to entry PF00106 adh_short,
Alcohol/other dehydrogenases, short chain type, score
-24.80, E-value 0.0011"
27323..28477
/gene="Cj1320"
27323..28477
/gene="Cj1320"
/feature="Cj1320, probable aminotransferase, len: 384 aa;
similar to members of the degt family e.g. TR:006953
(EMBL:X59554) Vibrio cholerae RBE protein (367 aa), fasta
scores: opt: 315 z-score: 377.4 E(): 1e-15, 32.7% identity
in 361 aa overlap. No Hp ortholog. Contains Pfam match to
entry PF01041 Degt_DnrJ_EryC1, Degt/DnrJ/EryC1/Stfs
family"
/codon_start=1
/transl_table=11
/product="putative aminotransferase (degt family)"
/protein_id="CAB73747.1"
/db_xref="GI:6968754"
/translation="MKKEKISFIRKSLFNKNENIALHEPCFIGNKKYLECIDSFVSS
VGEFVTRFEELAKERTKARFVATNMGRAILALANGIDNCGVITQISFVATN
ALAYGAKRVEDIDENLSTSPKALEHLENTQYOKDLSKGTTHKIKACVYNHT
RGLSHIKRIRKLCERYHILLDEBAKALGASTYENKALGYTRKCNINAAILFAGLENE
CGGALLSDENLAKLARHISTTAKIIPHEVYDHRIAVNRCINAAILFAGLENE
LLEKFRPVRKSLPSKAFQNCOSNELMWTKLEKRLINLSSVRIANKKE"
27389..28474
/gene="Cj1320"
/feature="Pfam match to entry PF01041 Degt_DnrJ_EryC1,
Degt/DnrJ/EryC1/Stfs family, score 245.90, E-value
5.7e-70"
28541..28550
/misc_difference
28541..28550
/feature="polymorphism G(10-11)"
28587..29129
/gene="Cj1321"
28587..29129
/gene="Cj1321"
/feature="Cj1321, probable transferase, len: 180 aa; similar
to TR:046674 (EMBL:U05248) E. coli polysialic acid gene
cluster region 2 neub (207 aa), fasta scores: opt: 267
z-score: 337.2 E(): 1.8e-11, 29.6% identity in 162 aa
overlap, some similarity to e.g. LPXA_BRUB
acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine
0-acyltransferase (283 aa), fasta scores: opt: 134
z-score: 173.4 E(): 0.025, 31.6% identity in 95 aa
overlap. No Hp match. Also similar to Cj1123c waiI (30.2%
identity in 179 aa overlap). Contains 2x Pfam match to
entry PF00132 hexapep. Bacterial transferase hexapeptide
(four repeats). Upstream region contains a G(10-11)
hypervariable tract"
/codon_start=1
/transl_table=11
/product="putative transferase"
/protein_id="CAB73748.1"
/db_xref="GI:6968755"

```

```

misc_difference
18035..18043
/Note="polymorphism C(9-10)"
gene
18701..19399
/Name="neuA2"
CDS
18701..19399
/Name="neuA2"
/EC_number="2.7.7.43"
/Note="Cj1311, neuA2, probable acylneuraminate
cytidyllyltransferase (CMP-N-acetylneuraminic acid
synthetase), len: 232 aa; similar to e.g. NEUA_ECOLI
acylneuraminate cytidyllyltransferase (EC 2.7.7.43)
(CMP-N-acetylneuraminic acid synthetase) (419 aa), fasta
scores: opt: 316 z-score: 379.1 E(): 8.2e-14, 29.0%
identity in 231 aa overlap, 39.6% identity to HP0326. Also
similar to Cj1143 (31.6% identity in 215 aa overlap), and
Cj1331 (32.9% identity in 234 aa overlap)"
/codon_start=1
/transl_table=11
/product="acylneuraminate cytidyllyltransferase"
/protein_id="CAB73738.1"
/db_xref="GI:6968745"
/translation="MKNLCTIPANGSKRIPRKNIIDELGKPLISYIEMALNSGIFD
EVLVSDDEELIEVALKYGAKAPVRKNIISDDVASSVAVQNAIEILSONQIYDH
CCLATAPALNKDLIKAYEKFIQNSKFLAEFEPIORAFYLNENNQVYDEK
HYKRSODLTAKAYADGAFYGTSGKAMLEDFIRPHSSVFLRNLCVDITDIDLE
FAKLYKVNHSAP"
19383..20207
/Name="Cj1312"
CDS
19383..20207
/Name="Cj1312"
/Note="Cj1312, possible flagellar protein, len: 274 aa;
similar to e.g. SPBG_BACSU spore coat polysaccharide
biosynthesis protein SPBG (222 aa), fasta scores: opt: 166
z-score: 208.1 E(): 0.00028, 24.8% identity in 226 aa
overlap, and TR:Q45987 (EMBL:U27302) Caulobacter
crescentus FLAR (329 aa), fasta scores: opt: 234 z-score:
286.7 E(): 1.2e-08, 23.1% identity in 277 aa overlap"
/codon_start=1
/transl_table=11
/product="possible flagellar protein"
/protein_id="CAB73739.1"
/db_xref="GI:6968746"
/translation="MKVLFPSDSSSOIGFGIKRDLVAKOYSPVAFCLPIESLID
EIPYVELESSESYELINLKEKPELLIIDHGISVDDEKLKLTGVKILISFDE
IKPHHCIDILNVAAYAKASDYEGVLPKCEVRCGFSTALIREEYQDAKNNREKKYDF
PICMGTDINLSQIASLSELPKTIISIASSSPNPKLQKQFAKLNHNLRLFDHNN
IAKLMSNKNLIISSASLVNEALLKANFKAICYVKNQESTATWLAKGYEVEKY"
20204..20677
/Name="Cj1313"
CDS
20204..20677
/Name="Cj1313"
/Note="Cj1313, possible flagellar protein, len: 157 aa;
similar to e.g. TR:Q45990 (EMBL:U28867) Caulobacter
crescentus FLAG. 31.9% identity to HP0327"
/codon_start=1
/transl_table=11
/product="possible flagellar protein"
/protein_id="CAB73740.1"
/db_xref="GI:6968747"
/translation="MILKNFAELNSQEIILFKRRNHPDISOFMKTKHIDFEHLRF
IRNLHDSNKKYFLVPODOIIGVIDVNTTKSCERGLATPDLKGVGVNLNETKK
YAFELIKVDFLTKAVYVFRDNHAKALKLYQGNHFTIYDEKDPYVYVCLQSHCKALPS"
20222..20605
/Name="Cj1313"
misc_feature
/Note="Pfam match to entry PF00583 Acetyltransf,
Acetyltransferase (GNAT) family, score 40.60, E-value

```

```

gene
3..5e-08"
complement(20642..21388)
/Name="Cj1314c"
CDS
complement(20642..21388)
/Name="Cj1314c"
/Note="Cj1314c, probable cyclase, len: 248 aa; similar to
many e.g. HIS6_BACSU HisF protein (cyclase) (252 aa),
fasta scores: opt: 562 z-score: 655.9 E(): 3.1e-29, 38.3%
identity in 253 aa overlap. No Hp match. Also similar to
hisF_Cj1603 (36.5% identity in 255 aa overlap). Contains
Pfam match to entry PF00977 His_biosynth, Histidine
biosynthesis protein"
/codon_start=1
/transl_table=11
/product="putative cyclase"
/protein_id="CAB73741.1"
/db_xref="GI:6968748"
/translation="MLKRRILPCVILKNSQVLYKSEFKDFRTIGLHSTMRITNARNV
DELLIDLDASKSGTIDRESIEDLAKCEPMLPTGGIKTLEDIOKLINTLQADKTSIN
SKALDMDPISKAGNRFSGOCISIDVRRGDGFCVYDRGNLSEKSPLEALEYERK
GAGELLHSDPEEGKAKGYDLELKIPONKLKPLILINGLSKPSDGEALNLGADAL
AGAYIFHPSKYTPKDVKEELAROGFAVRLL"
complement(20675..21379)
/Name="Cj1314c"
misc_feature
/Note="Pfam match to entry PF00977 His_biosynth, Histidine
biosynthesis protein, score 205.30, E-value 9e-58"
complement(21389..21994)
/Name="Cj1315c"
CDS
complement(21389..21994)
/Name="Cj1315c"
/Note="Cj1315c, probable amidotransferase, len: 201 aa;
similar to many e.g. HIS5_BACSU amidotransferase HIS5 (212
aa), fasta scores: opt: 536 z-score: 660.1 E(): 1.8e-29,
41.3% identity in 208 aa overlap. No Hp match. Also
similar to hisG_Cj1600 (39.2% identity in 199 aa overlap).
Contains PS00442 Glutamine amidotransferases class-I
active site, and Pfam match to entry PF00117 GATase,
glutamine amidotransferases class-I"
/codon_start=1
/transl_table=11
/product="amidotransferase"
/protein_id="CAB73742.1"
/db_xref="GI:6968749"
/translation="WALIDYKAGNLNSVAKAFKIGAINFIANKPNPLQKADLLP
GVGSFKAMKMLKEGLFEALKEQVILVOKKPIILGICGMOLFEGYEGVCEGLGTI
EGEVVKEEEDLNKILPHGMNELELIGLOVLYGIDDKSDPYFHSYVYVCKDFPSVA
KAOYGHKFVASLQKDHIFATQFHEKQNLGKLLENFARL"
complement(21392..21994)
/Name="Cj1315c"
misc_feature
/Note="Pfam match to entry PF00117 GATase, glutamine
amidotransferases class-I, score 13.80, E-value 2e-07"
complement(21737..21772)
/Name="Cj1315c"
misc_feature
/Note="PS00442 Glutamine amidotransferases class-I active
site"
complement(21991..23127)
/Name="Cj1316c"
CDS
complement(21991..23127)
/Name="Cj1316c"
/Note="Cj1316c, unknown, len: 378 aa; similar to e.g.
TR:P72140 (EMBL:050396) Pseudomonas aeruginosa WbpG
(11p)polysaccharide gene cluster) (376 aa), fasta scores:
opt: 257 z-score: 302.0 E(): 1.6e-09, 23.5% identity in
366 aa overlap. No Hp match. Also similar to Cj1324 (33.2%
identity in 373 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1316c"
/protein_id="CAB73743.1"
/db_xref="GI:6968750"
/translation="MKFCKKCVMPDKPDLHPDEGVCACRSQAKNNQINMQEKEK
EPEELIKKKYKHPYDCVIGCSGKRSSTQYVKNMELGMLPDCVCEPSVPTIKGRKN
LDNLNHLGVDLTHIKRDPKYQKLAREAPLRTIGDNOMNHLGTFSTPRAVNFVPL

```

```

gene      KNLFITKITAEDLKNCKVSDLINLAKLD"
complement(12027. .13244)
/gene="Cj1305c"
complement(12027. .13244)
/gene="Cj1305c"
/note="Cj1305c, unknown, len: 405 aa; no Hp match. A
member of the 617 family of C.j. proteins containing
homopolymeric tracts; similar to Cj0617/Cj0618 (38.9%
identity in 180 aa overlap/37.1% identity in 221 aa
overlap), Cj1306c (78.1% identity in 407 aa overlap),
Cj1310c (63.7% identity in 402 aa overlap), Cj1342c (38.0%
identity in 426 aa overlap). Contains a variable
polyg-tract at aa 194; G(9) (consensus) gives this ORF,
G(10) would cause truncation after 22 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1305c (617 family)"
/protein_id="CAB73732.1"
/db_xref="GI:6968739"
/translacion="MLINQTEIDSCDVELINIKRTSKLEYRSYDDEKEIKAVFII
GGYGANAIYFELDSYRNYIAKNPDVAVHVFYHCFQCRSDVEKYSAYKFEEDIN
IKNLNHFHSYSGEINRNDNLFLANSIVKHVENLKMGNKLDHNFKNLFTSPFIPEDN
KONGMAADHINAKDLVKRPKPADLPRIYGGSGYGLSLIKAKIPMYDGYI
DNGSALPLPLNYIDGEMHSYGDYEDFPHNRITFLKTHWTKENSYPFENNENF
IRTLNKLHLILLOSQKNKNIIVSYSDKDPLEPANRQOTMOLIKILGDSVSLNLD
ENKIDGKFINLHDGCGIPDKALFKELPLMLEKIQKRSFMOENSISYCGNKVFTE
KVDGDKFELVIRK"
misc_difference 12657. .12665
/note="polymorphism C(9-10)"
complement(13257. .14483)
/gene="Cj1306c"
complement(13257. .14483)
/gene="Cj1306c"
/note="Cj1306c, unknown, len: 405 aa; no Hp match. A
member of the 617 family of C.j. proteins containing
homopolymeric tracts; similar to Cj1310c (59.8% identity
in 410 aa overlap), Cj1305c (78.1% identity in 407 aa
overlap), Cj1342c (38.5% identity in 426 aa overlap) and
Cj0617/Cj0618 (39.2% identity in 181 aa overlap/39.5%
identity in 220 aa overlap). Contains a variable
polyg-tract at aa 194; G(9) (consensus) gives this ORF,
G(8) would cause truncation after 5 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1306c (617 family)"
/protein_id="CAB73733.1"
/db_xref="GI:6968740"
/translacion="MLINQSFETIDSCDVELINIKRTSKLEYRSYDDEKEIKAVFII
GGYGANAIYFELDSYRNYIAKNPDVAVHVFYHCFQCRSDVEKYSAYKFEEDIN
IKNLNHFHSYSGEINRNDNLFLANSIVKHVENLKMGNKLDHNFKNLFTSPFIPEDN
KONGMAADHINAKDLVKRPKPADLPRIYGGSGYGLSLIKAKIPMYDGYI
DNGSALPLPLNYIDGEMHSYGDYEDFPHNRITFLKTHWTKENSYPFENNENF
IRTLNKLHLILLOSQKNKNIIVSYSDKDPLEPANRQOTMOLIKILGDSVSLNLD
ENKIDGKFINLHDGCGIPDKALFKELPLMLEKIQKRSFMOENSISYCGNKVFTE
KVDGDKFELVIRK"
misc_difference 13896. .13904
/note="polymorphism C(8-9)"
14544. .16052
/gene="Cj1307"
14544. .16052
/gene="Cj1307"
/note="Cj1307, possible amino acid activating enzyme, len:
502 aa; similar to e.g. DTRA_BACSU D-alanine-activating
enzyme (EC 6.3.2.-) (503 aa), fasta scores: opt: 726
z-score: 837.7 E(1): 0, 30.3% identity in 495 aa overlap,
and to parts of ENRF_ECOLI enterobactin synthetase
component F (1293 aa), fasta scores: opt: 549 z-score:
627.8 E(1): 1.1e-27, 27.4% identity in 460 aa overlap, and
GRSB_BACBR gramacidin s synthetase II (4451 aa), fasta
scores: opt: 571 z-score: 644.7 E(1): 1.3e-28, 29.9%
identity in 508 aa overlap. No Hp match. Contains PS00455
Putative AMP-binding domain signature, and Pfam match to
entry PF00501 AMP-binding, AMP-binding enzyme"
/codon_start=1
/transl_table=11
/product="putative amino acid activating enzyme"
/protein_id="CAB73734.1"
/db_xref="GI:6968741"
/translacion="MTTHITDFLEKSLIKSEKTAPEVPYKKEKITTYDFDLPSK
LASEILKTLDNDNTQAPVILIPKIGDCLISFVALSGNFYLLDLDEKSPREVEKY
IEVLPKLFITSKDNLNLEPLTYQDFPSFNDESLIHAKKHIDTNLLYFETS
GSTGIPKGSVAHKSVIDYEFWCEFPDENETLANOAPPEPNSIIDFSSVSGA
TILHLPNHLFAFPKLTIECEKEKVSITFPVSYLYEATNENLNSNAIKRMFLC
GEIHPNKOLINIKRHLPLNALFANILGTEITDVCSEFIINRTFDESLIPGKACKT
ELLVFDENMNFISPKQIGVGELEFRGTSLSLGVNDKETKQAFIONPLHNDLIL
KYGEDIVSYNEFGLICGRADNOIKMGHRIELGEIESVINSHVNTKNSACIFKEDI
ICEFSESEINFPKDLKPLSPWIPKXFIKDKFKLNQNGKIDRKVILELY"
14643. .15875
/gene="Cj1307"
/note="Pfam match to entry PF00501 AMP-binding,
AMP-binding enzyme, score 218.90, E-value 7.4e-62"
15009. .15044
/gene="Cj1307"
/note="PS00455 Putative AMP-binding domain signature"
16092. .16319
/gene="acpp4"
16092. .16319
/gene="acpp4"
/note="Cj1308, acpp4, possible acyl carrier protein, len:
75 aa; some similarity to e.g. ACP_CYP4 acyl carrier
protein (103 aa), fasta scores: opt: 84 z-score: 126.5
E(1): 9.5, 31.3% identity in 64 aa overlap. No Hp match"
/codon_start=1
/transl_table=11
/product="putative acyl carrier protein"
/protein_id="CAB73735.1"
/db_xref="GI:6968742"
/translacion="MQEIKQFPINIEFTDINESKENLVSEDFDIDSLMALVAETEF
YKRLKAEFTTPENFESFENIKMLFAMKD"
complement(16316. .17386)
/gene="Cj1309c"
complement(16316. .17386)
/gene="Cj1309c"
/note="Cj1309c, unknown, len 356 aa; no Hp match"
/note="Cj1309c, unknown, len 356 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1309c"
/protein_id="CAB73736.1"
/db_xref="GI:6968743"
/translacion="MTAVLHIGTTNGNCKQGLFMQNEKLLQAKIYIKSKLRVNR
HMAVLVAVLELVOKEDILKREVSLSHTTNERLLRAIENFSESALHDKKFIYSAEI
VWPFSTKXHEIIEKIRLEGFTQIYIIVFRDPLAGLNSGSDHNNNGYSADFA
POHPRKYIYDYMICKTHAEVGEENLIVRLREYVGGLTKDFYHGLGMDSEF
VLKOTKRESNPLGMLMSRLNCKDKODNLNSLFLPARRPFESKREKRLFPVOKDI
AKAVDYFASLSLGMVKNKRYPRHKNLSLETPVNMWEYEQNTLTLMLSKMDVDVAFINQ
ITVSKNETISSLEQDELRKRD"
complement(17408. .18622)
/gene="Cj1310c"
complement(17408. .18622)
/gene="Cj1310c"
/note="Cj1310c, unknown, len: 404 aa; no Hp match. A
member of the 617 family of C.j. proteins containing
homopolymeric tracts; similar to Cj1305c (63.7% identity
in 402 aa overlap), Cj1306c (59.8% identity in 410 aa
overlap), Cj1342c (37.1% identity in 426 aa overlap) and
Cj0617/Cj0618 (37.4% identity in 182 aa overlap/34.1%
identity in 220 aa overlap). Contains a variable
polyg-tract at aa 194; G(9) (consensus) gives this ORF,
G(10) would cause truncation after 22 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1310c (617 family)"
/protein_id="CAB73737.1"
/db_xref="GI:6968744"
/translacion="MLINQTEIDSCDVELIGIKRTSKLEYRSYDDEKEIKAVFII

```

overlap). Similarity continues from upstream gene Cj1296. There is a G(9) tract upstream of this start that, if variable, would allow translation from Cj1296"

/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1297"
/protein_id="CAB73724.1"
/db_xref="GI:6968731"
/translation="MKGAKADIFLKNSSCGKGSVEILTRREGCKMLGLNGHAL
THIAEINTSCRYFKEKGFVIDELGQKREKRYVVRDLKAYVCSLDKINIVRO
TRYKSIKIFAGDFLESDAKYKVAIGNAKQDKFAIYEKILL"
6831..7622
/gene="Cj1298"
6831..7622
/gene="Cj1298"
/note="Cj1298, unknown, len: 263 aa; weak similarity to aminoglycoside N³-acetyltransferases e.g. AAC4, SALSP aminoglycoside N³-acetyltransferase IV (261 aa), fasta scores: opt: 139 z-score: 177.8 E(): 0.014, 25.7% identity in 175 aa overlap. Similar in N- and C-terminus to upstream ORFs Cj1296 (65.7% identity in 105 aa overlap), and Cj1297 (44.7% identity in 141 aa overlap). No Hp match. There is a G(9) tract that, if variable, would allow fusion of Cj1296 and Cj1297. There is no equivalent repetitive tract in this CDS"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1298"
/protein_id="CAB73725.1"
/db_xref="GI:6968732"
/translation="MKYLEFEENKKYSYNDPIETEFYKLGQKDTLCVHLEENGGP
LISNREFLOTIDCFEVIKKEGTLPTSPYSCKNKENVYDKINSKTRKMLNFK
OGYKRNTPDIPFAIKGAKKELELKDTTSCGNCYEVLTENGYMTFGGGHTL
THAEOPNVFTRHKIFSGILDENNI SYKNTSYVRLKIDISSEPLNIINIIVNN
TKNFKNNFAGDHNINAKYEIILMKKIDINOTILIENVTIY"
7653..7883
/gene="acpP2"
7653..7883
/gene="acpP2"
/note="Cj1299, acpP2, probable acyl carrier protein, len: 76 aa; similar to many e.g. ACPH, MYCGE acyl carrier protein homolog (84 aa), fasta scores: opt: 111 z-score: 172.6 E(): 0.026, 35.8% identity in 81 aa overlap. No Hp match. Also similar to acpP Cj0441 (34.1% identity in 41 aa overlap)"
/codon_start=1
/transl_table=1
/product="putative acyl carrier protein"
/protein_id="CAB73726.1"
/db_xref="GI:6968733"
/translation="MEKQFYELLENILETVVDENTNLSMENCKMWSLAHIDIIMSL
EERIKNKEDLSLKSQSALEKIOTLKAKE"
7880..8773
/gene="Cj1300"
7880..8773
/gene="Cj1300"
/note="Cj1300, unknown, len: 297 aa; no Hp match"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1300"
/protein_id="CAB73727.1"
/db_xref="GI:6968734"
/translation="MKENIKELFOALNQKPHISKYIKTIEFSQDEEERDKLISYII
QDGISIKQDCYLLINDYLEETKEFTEINGAVRSSDEVKDKVFNESYMAQYMG
LALSFWMAIHVIRKYSQYIQNPTPKTYFEIGPQHGSEFARAVYSGKFSFAGLD
ISPTSCELTGKMWKQOYGLITKCDPFTYDPDKRADILVIGVLEHVKPLEEF
LKRSEKLSGSGEFPATIPINAPADIDHIVLYFSHPDEVKDLIEKSGLKFKAYCEPMAND
YSLEKALKENATMAVVLNA"
8766..9182
/gene="Cj1301"
8766..9182
/gene="Cj1301"
/note="Cj1301, unknown, len: 138 aa; similar to a

hypothetical protein from *Pyrococcus horikoshii* TR-058010 (EMBL:AP000001) PH0272 (136 aa), fasta scores: opt: 137 z-score: 188.8 E(): 0.0032, 32.8% identity in 137 aa overlap. No Hp match"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1301"
/protein_id="CAB73728.1"
/db_xref="GI:6968735"
/translation="MLNTTLPLPHHIGVACKNLEKEKEEFKIGFYKESPEIDKOG
VREGIIPICNEAPFOYRELLQNLNDGPDLSYKLNKTKMHLVYESKNINADLILLE
OOGGICIVPIWQASVFAKLCITMMPNLLITLVEIK"
9179..10741
/gene="Cj1302"
9179..10741
/gene="Cj1302"
/note="Cj1302, unknown, len: 520 aa; no Hp match"
/codon_start=1
/transl_table=1
/product="hypothetical protein Cj1302"
/protein_id="CAB73729.1"
/db_xref="GI:6968736"
/translation="MNLSPFLKRNELIKFAKEIDFSKSDLLINVRNNAPEVOGI
IAPILHRANLAEENFSSYDLSIFDNFKKESLELMLDTRKANNQNFTEERLEL
RKISQNPILVLSGEFTDKILNCEIFNIEKLKEFDEEDJIDLAKBELTSGKLN
KALIFLQILGLSLIPALVPKALVALYDLNDLYOGIIGEGIDNLSPLKTLQD
KIKTFKQGFELALASNEEKDARKLEIFRIDMDGDFDAPRANINPEKENTLKE
KFINNTNAMFLIDNIAELENTKITGKTLGCENLTKLEPLNLIKNTKEDOT
RAKDIAANALBEELKSLSDSEYFQNLRETSNEDSNDLONIPRISEILGKTQNPANT
RLNDEKVAQHOKRELYSVSSDLSGSIATVFSCKBSNLTDDICISCRNLGRK
LEIMFTRAEFLALHFEDLKNNNARLYQGERNMPLSFLEQISKBEFNSALIPRO
NLNFKFLIHEN"
10731..11792
/gene="fabH2"
10731..11792
/gene="fabH2"
/EC_number="2.3.1.41"
/note="Cj1303, fabH2, probable 3-oxoacyl-[acyl-carrier-protein] synthase, len: 353 aa; similar to e.g. FABH, SPIOL 3-oxoacyl-[acyl-carrier-protein] synthase III precursor (EC 2.3.1.41) (405 aa), fasta scores: opt: 418 z-score: 484.2 E(): 1.1e-19, 27.4% identity in 340 aa overlap. No Hp ortholog. Also similar to fabH Cj0328c (31.3% identity in 339 aa overlap)"
/codon_start=1
/transl_table=1
/product="putative 3-oxoacyl-[acyl-carrier-protein] synthase"
/protein_id="CAB73730.1"
/db_xref="GI:6968737"
/translation="MKTRPDKAKISGICVSVPEHKICIDDELSEFSNDIKTRMK
VIGINTRYICDENTCVSDIGKHAANTLQGINIDKNSLDLIVTOSRPSDPMSTAK
LHOLLNSSKTVAPDLSQACAGYIYGFVASHLSIQSLGKILILCGTTSKFTIPKRM
NLATPFGDVSATILKTDENAEFEIGSOGKFFDKLIRPGARIPKADIPNDOSLM
QTEEFQLEMLYMDGANIFNMALCEKSPREKLEFEKVBKEDIAFLFQSNAYLVD
CIKEELKINDKVPNFTMERVYANISASPLTLCEDTDPREFKASLSAFQAGLSWCSA
VLNFKDLYTKNIIITYTEK"
11792..12013
/gene="acpP3"
11792..12013
/gene="acpP3"
/note="Cj1304, acpP3, probable acyl carrier protein, len: 73 aa; similar to e.g. CP MYXA acyl carrier protein (78 aa), fasta scores: opt: 101 z-score: 157.0 E(): 0.19, 28.4% identity in 74 aa overlap. No Hp match. Also similar to acpP2 Cj1299 (25.4% identity in 63 aa overlap)"
/codon_start=1
/transl_table=1
/product="putative acyl carrier protein"
/protein_id="CAB73731.1"
/db_xref="GI:6968738"
/translation="WKTPELLEKEMHRDEDLNDMDLLDIDEMDSLAFVSVVLY

```

misc_feature
/feature="acch"
/note="Pfam match to entry PF00364 biotin_req_enzy,
Biotin-requiring enzymes, score 104.20, E-value 2.5e-27"
1974..2534
/gene="dcd"
1974..2534
/gene="dcd"
/EC-number="3.5.4.13"
/note="Cj1292, dcd, possible deoxycytidine triphosphate
deaminase, len: 186 aa; similar to many predicted dcd
proteins, and to DCD_ECOLI deoxycytidine triphosphate
deaminase (EC 3.5.4.13) (193 aa), fasta scores: opt: 160
z-score: 203.5 E(): 0.00049, 28.8% identity in 160 aa
overlap. 71.4% identity to HP0372. Contains Pfam match to
entry PF00692 dUTPase"
/codon_start=1
/transl_table=11
/product="possible deoxycytidine triphosphate deaminase"
/protein_id="CAB73719.1"
/db_xref="GI:6968726"
/translation="MGLADNMIRKMALEHKMIEPCEANIGKGVSYGLSYGDIR
VGRFKIETNVSTVDPKNFVENVDPEGDNCIYVANSFALARTIEYKMDNVIA
ICLGKSTYARCGILVNTPEPGECHITTEISNTTLPAAKITANEGIAOVLEQDGE
KCDITTKKRGKTYOAGTITLPRLK"
2046..2522
/gene="dcd"
/note="Pfam match to entry PF00692 dUTPase, dUTPase, score
3.90, E-value 0.0016"
2584..3588
/gene="Cj1293"
2584..3588
/gene="Cj1293"
/note="Cj1293, possible sugar nucleotide
epimerase/dehydratase, len: 334 aa; similar to many e.
TR:045984 (EMBL:027301) Caulobacter crescentus FLA1
protein (331 aa), fasta scores: opt: 1240 z-score: 1410.9
E(): 0.57.8% identity in 329 aa overlap, CAPD_STRAU CAPD
protein (599 aa), fasta scores: opt: 529 z-score: 602.6
E(): 2.9e-26, 33.9% identity in 319 aa overlap, and
TR:069130 (EMBL:AF064070) Burkholderia pseudomallei
putative epimerase/dehydratase WB1 (637 aa), fasta
scores: opt: 433 z-score: 493.6 E(): 3.4e-20, 30.5%
identity in 282 aa overlap. 64.6% identity to HP0840.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="possible sugar nucleotide epimerase/dehydratase"
/protein_id="CAB73720.1"
/db_xref="GI:6968727"
/translation="MENKNKILITGGSGFKTYTKVLENNYKPKLIISRDELKOF
EMASVFNAPCMRYEIGVDKRLSAMRDVFIHAAMKHVIAEYVNMEDIKRI
HGQNVITDACFENGKCCILSTDKACNPNVIGATRLASDKLFVANNINAKGKQRF
GVTIRGNVGSRGSVVPEFKKLISEGAKELPITDTRTRFWISLEDDGKRVLSNFERM
HGEISFIPKIPSMKITDLAHLAPNLNLSKLTIGRAGKXLEHIMISSDSHLLTFEYFY
YAISSPIKFPVKDNDSEINALEGKGKVKDGSYSDNNPLMASSEKLEIINHTEGF
"
2617..2640
/gene="Cj1293"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
3590..4720
/gene="Cj1294"
3590..4720
/gene="Cj1294"
/note="Cj1294, probable aminotransferase, len: 376 aa;
similar to many members of the degT family e.g. SSSC_BACSU
spore coat polysaccharide biosynthesis protein (389 aa),
fasta scores: opt: 586 z-score: 676.7 E(): 2.1e-30, 35.3%

```

```

misc_feature
/feature="putative aminotransferase (degT family)"
/product="putative aminotransferase (degT family)"
/protein_id="CAB73721.1"
/db_xref="GI:6968728"
/translation="MLTYSHQNIQSDIDITLFTALKDELITGGKKNVEFEALCEYNG
VKHACVNSATSLHLAVTAQVEKIVLTPTLFATANAALMAGAKVEFIDIKNG
NIBERKLEARLKESENIQAISVYDPAVNSVDEDEISLTKKYNIPLIDDSHALGAL
YKSEKVGKRAKDLSIESFHPVKPIITPEGGAIVSDNEELIDKILRSHGIKKRIMWS
DMVELGNYRLSDVACALGINOLAKLDHNEKEBEIANFPDKEREKPYSTIKINY
KSSRHLYPLILPPEFYCQKEELFESLHAGIGVQVHKPTPEFSFYKILGELIKON
ADNFRKALSLIPCHOENMLKDAFVKDTLFSILEKVKGYCG"
3608..4681
/gene="Cj1294"
/note="Pfam match to entry PF01041 DegT_DnrJ_EryC1,
DegT/DnrJ/EryC1/Stis family, score 374.40, E-value
1.2e-108"
4713..6020
/gene="Cj1295"
4713..6020
/gene="Cj1295"
/note="Cj1295, unknown, len: 435 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1295"
/protein_id="CAB73722.1"
/db_xref="GI:6968729"
/translation="MDKLDFRSQDFSGTGAETLACELFPPIRSITGCGFRDSELT
NKTLLGGILKFNHISKGTQYFVDIVPEMNVKAYITTPGKERICDEKKNLHLNLS
EAIQDEIELELDHLYSIEEMDAPVYVYKRRGFCITNERRKLLKGGKYKYI
DAKHDENGVLDVADFEILPSTONKDELITSTYCHPSMANNEISGPAVSIFLAKMLG
LKRERNYRPRVILPBTIGSYLSKILFCLIKRKGKGVASCIGDDHTSYLHTFVGN
TISDKVALHTLKKENKFAFSPLDRSDENQYNAPLVNIAGIVCYCRRYRDYGIANS
KQDNLFISEKGGQSLQSMOEMTLNLEINNVYNTIVCEPNLRGLYHPLSTNDIP
LACNFAYCGDGDIDIANILNMOAYEFKELLEKIFGVLR"
4856..4864
/note="G(9)"
6017..6376
/gene="Cj1296"
6017..6376
/gene="Cj1296"
/note="Cj1296, unknown, len: 119 aa; weak similarity to
aminoglycoside N3'-acetyltransferase e.g. AAC3_PSEAE
aminoglycoside N3'-acetyltransferase III (271 aa), fasta
scores: opt: 102 z-score: 144.2 E(): 1.31.7% identity in
60 aa overlap. No Hp match. Similar to N-terminus of
Cj1298 (65.7% identity in 105 aa overlap). Similarity
continues in downstream ORF Cj1297. There is a G(9) tract
at aa 104 that, if variable, would allow translation to
continue into Cj1297"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1296"
/protein_id="CAB73723.1"
/db_xref="GI:6968730"
/translation="MKYFLHNKGRKYSDDKLDIAFYOLIGKRDITLVHTELMKPKKA
LTKNDPLKTLCCFFRVLAGKEGTLMLPTFTYSFCXNEYVDKXHGKGVLENEPFT
SGGGRUGELATFFPL"
6326..6334
/note="G(9)"
6373..6810
/gene="Cj1297"
6373..6810
/gene="Cj1297"
/note="Cj1297, unknown, len: 145 aa; no Hp match. Similar
to C-terminus of Cj1298 (44.7% identity in 141 aa

```

```

SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 431)
AUTHORS     Nano, F.E.
TITLE       Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL     Patent: US 6228371-A 111 08-MAY-2001;
FEATURES
    source   1..431
            Location/Qualifiers
BASE COUNT  58 a 122 c 154 g 92 t 5 others
ORIGIN
AR149066 Length: 431 February 21, 2002 16:59 Type: N Check: 1540
Found using 'seq2-3' (pappu03.key)

1      TGTCCCGCATGTAGTCGGCTGCNCGTATCGCTTGACAGCTTTNGCCGTGATGTG
            |-----|
            33      38
61      AGAAGAATATGTTGGTATCACCATTCTTCTGCTGCGCGGATGACGCGCT
            |-----|
            115 120
121     GCGGTCGACGANGTCGGGATGTCACCGTGAATACGCCATCGGTACCATCGCGGGG
181     CTGCTTCGGTGCATCTCTACACGGTCTGCACCGGGGATTCATTGTGTGCGCTCA
            |-----|
            207 212
            233 238
241     ACCGCATCATCGTGTGCGCGCTCAGCACAGTTTACGCTGCTGGGGTGGAGAC
            |-----|
            277 282
301     CGTGAAGCGCGCTTGGCGATGCCACCTGCTGCTGCTGCTGCTGCTGCTGCGCGG
            |-----|
            310 315 320 325
            316 321
361     ---|
            364
            CGTACCGCGGTATCAATGCAAGTGGCTGTATGACGCGGCCGCGAGCGCCG
            364

-----
734 matches found in sequence:
cj11168x5 ; TOIG of: cj11168x5 check: 8802 from: 1 to: 263335
(from "mycobacter-se.seq")
TOIG of: cj11168x5 check: 8802 from: 1 to: 263335

LOCUS      CJ11168x5 263335 bp DNA BCT 08-JUL-2000
DEFINITION Campylobacter jejuni NCTC11168 complete genome; segment 5/6.
ACCESSION AL139078 AL111168
VERSION    AL139078.2 GI:6968723
KEYWORDS
SOURCE     Campylobacter jejuni.
ORGANISM   Campylobacter jejuni
            Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
            Campylobacter.
REFERENCE  1 (bases 1 to 263335)
AUTHORS     Parkhill, J., Wren, B.W., Mungall, K., Kelley, J.M., Churcher, C.,
            Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holtroyd, S.,
            Jagels, K., Karlyshev, A., Moule, S., Pallen, M.J., Penn, C.W.,
            Quail, M., Rajandream, M.A., Rutherford, K.M., VanVleet, A.,
            Whitehead, S. and Barrett, B.G.
TITLE       The genome sequence of the food-borne pathogen Campylobacter jejuni
            reveals hypervariable sequences
JOURNAL     Nature 403 (6770), 665-668 (2000)
MEDLINE    20150912
REFERENCE   2 (bases 1 to 263335)

```

```

AUTHORS     Parkhill, J.
TITLE       Direct Submission
JOURNAL     Submitted (09-FEB-2000) Submitted on behalf of the Campylobacter
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of C. jejuni sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/C-jejuni/).
FEATURES
    source   1..263335
            Location/Qualifiers
1..263335
/organism="Campylobacter jejuni"
/strain="NCTC 11168"
/db_xref="taxon:197"
complement(38..1369)
/gene="accC"
complement(38..1369)
/gene="accC"
complement(47..1357)
/gene="accC"
/EC_number="6.4.1.2"
/EC_number="6.3.4.14"
/note="Cj1290C, accC, probable biotin carboxylase (subunit
of acetyl-CoA carboxylase (EC 6.4.1.2)), len: 443 aa;
highly similar to many e.g. ACCC-ECOLI biotin carboxylase
(EC 6.3.4.14) (449 aa), fasta scores: opt: 1560 z-score:
1741.0 E(): 0.53.2% identity in 442 aa overlap. 67.0%
identity to HP0370. Also similar to Cj1037C pyca (46.5%
identity in 445 aa overlap). Contains PS00866 and PS00867
and Pfam match to entry PF00289 CPSase_L_chain,
Carbamoyl-phosphate synthase (CPSase)"
/codon_start=1
/transl_table=11
/product="biotin carboxylase"
/protein_id="CA873717.1"
/db_xref="GI:6968724"
/translation="METKSLILANRGRIALRLTIEMGKKAICVSEADKALYIK
YADASTICGARSESELTIPALIAAEIAEADAFPGYFLSENFVICAHNK
FIGPSVEMAMIMSDSKAKOVMOAGVPIPGSGALAGAFAAKKLAKEIGVYLKA
AAGGGGRMVRVNEKDLERKAYSAEAMTAFGDMYMEKYIONRHIIEVYIGS
FGNVIHVGEDCSMORRHOKLIESPAILLDETRTRTHETRAKAAIGYEGNGFS
FLVDKNDLPEFIEMNTRLOVHCYSEVSGIDIIEMIKVAEGVALPSQSLTNGS
IECRITAEDESKTFLPSPKITTKYIIPRAGRVRNESHQYQSYPAYIDSITGLVWA
EDRNKAIAKKVALDELLIGITTKDFHLSKMNPDFINNNYDTYLARH"
complement(47..1357)
/gene="accC"
/note="Pfam match to entry PF00289 CPSase_L_chain,
Carbamoyl-phosphate synthase (CPSase), score 640.70,
E-value 8e-189"
complement(488..511)
/gene="accC"
/note="PS00867 Carbamoyl-phosphate synthase subdomain
signature 2"
complement(866..910)
/gene="accC"
/note="PS00866 Carbamoyl-phosphate synthase subdomain
signature 1"
complement(1371..1826)
/gene="accB"
complement(1371..1826)
/gene="accB"
/EC_number="6.4.1.2"
/note="Cj1291C, accB, probable biotin carboxyl carrier
protein of acetyl-CoA carboxylase, len: 151 aa; similar to
e.g. BCCP-ECOLI biotin carboxyl carrier protein of
acetyl-CoA carboxylase (EC 6.4.1.2) (156 aa), fasta
scores: opt: 307 z-score: 329.6 E(): 4.6e-11, 38.9%
identity in 162 aa overlap. 41.4% identity to HP0371.
Contains Pfam match to entry PF00364 biotin_reg_enz,
Biotin-regulating enzymes"
/codon_start=1
/transl_table=11
/product="putative biotin carboxyl carrier protein of
acetyl-CoA carboxylase"

```



```

...
188  ATGGTTGGGAGTAGATCTCCAGTCGCCACTGATATAACTGGGTGATGTCGTGTGA      |-----|
248  GCGGACAGGATAGAGGCGCATGACCG                                     238  243

-----
6 matches found in sequence:
ar149057 ; TOIG of: ar149057 check: 6558 from: 1 to: 252
(from "mycobacter-ge.seq")
TOIG of: ar149057 check: 6558 from: 1 to: 252

LOCUS      ARI49057      252 bp      DNA
DEFINITION Sequence 73 from patent US 6228371.
ACCESSION  ARI49057
VERSION    ARI49057.1 GI:15113648
KEYWORDS
SOURCE     .
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 252)
AUTHORS    Nano,F.E.
TITLE      Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL    Patent: US 6228371-A 73 08-MAY-2001;
FEATURES   Location/Qualifiers
            source          1..252
                        /organism="unknown"
BASE COUNT  31 a      74 c      88 g      59 t
ORIGIN
ARI49057 Length: 252 February 21, 2002 16:59 Type: N Check: 6558
Found using 'seq2-3' (pappu403.key)

1  GTGATGCTTCACAGATTGATGTGCTCGGTTGATGCTGTGGCGACAGATAAACCGC      |-----|
                        25 30      44 49
61  CTGTTGGGGTGCGTGGCTCTGTGCTGGCAGCGACACTGCTCAAGCGCGCTGTGGCTGCT      |-----|
                        102 107
121  GTGATGCTGGCCAGTCGTGTGGCAGTGGGTTTCACGCTGGGCGTACTTCACGCGCTT      |-----|
                        175 180
181  CCTGCTGCTACAGGTGGCGGCTCAACGCGTGTGCGCGCATCGATATCGTGATCAGCGT      |-----|
                        212 217
241  CCTGCGCGCCGGA
241
-----
3 matches found in sequence:
ar149058 ; TOIG of: ar149058 check: 843 from: 1 to: 160
(from "mycobacter-ge.seq")
TOIG of: ar149058 check: 843 from: 1 to: 160

LOCUS      ARI49058      160 bp      DNA
DEFINITION Sequence 74 from patent US 6228371.
ACCESSION  ARI49058
VERSION    ARI49058.1 GI:15113649
KEYWORDS
SOURCE     .
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 160)
AUTHORS    Nano,F.E.

```

```

TITLE      Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL    Patent: US 6228371-A 74 08-MAY-2001;
FEATURES   Location/Qualifiers
            source          1..160
                        /organism="unknown"
BASE COUNT  14 a      57 c      68 g      21 t
ORIGIN
ARI49058 Length: 160 February 21, 2002 16:59 Type: N Check: 843
Found using 'seq2-3' (pappu403.key)

1  GCGCGCGCGCTGCTGCTGCGCCCGCGCGGGTGGGGTGCCGACGCTGTTCCGCC      |-----|
                        1 6
61  ACTGCGCGCGCGGACGATTTGGCGCGCTCTCGACGACGACAGACGAGCGTGGGGGC      |-----|
                        86 91
121  GCGGCTGCTGCGCGCGCGCTGGCCAGCTGCCACCGCGGTGG                |-----|
                        104 109

-----
8 matches found in sequence:
ar149059 ; TOIG of: ar149059 check: 474 from: 1 to: 401
(from "mycobacter-ge.seq")
TOIG of: ar149059 check: 474 from: 1 to: 401

LOCUS      ARI49059      401 bp      DNA
DEFINITION Sequence 75 from patent US 6228371.
ACCESSION  ARI49059
VERSION    ARI49059.1 GI:15113650
KEYWORDS
SOURCE     .
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 401)
AUTHORS    Nano,F.E.
TITLE      Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL    Patent: US 6228371-A 75 08-MAY-2001;
FEATURES   Location/Qualifiers
            source          1..401
                        /organism="unknown"
BASE COUNT  68 a      120 c      119 g      80 t      14 others
ORIGIN
ARI49059 Length: 401 February 21, 2002 16:59 Type: N Check: 474
Found using 'seq2-3' (pappu403.key)

1  ATCAGCCGCGGGTGAGCGCCCGCGATGACCTCGACGTGCTGCTGCGTCCGCTACTC      |-----|
                        15 20      33 38 42 47
61  AATCAATCACCATCTCTTTACGCACTTCTTAGAGATGTG                |-----|
                        36 39 41 45 50

...
131  TTGCGGCTCGCGGTTGTGATGATCTTGGCCACAGACTTGATGTCTTAACGACACTA      |-----|
191  GTCCGTCGGGACNGCAANCCCGCACCGTGGAGTGTGCTGACG                |-----|
181  186

```

345 |
CGGGCGGCTGGGTACCGAGCTGGACACCGAAAGGCCGAGTGCASAAAGCATGAGACA |-----|
345 391 396
405 GCTGGCGAGAGACTACTGCTGGCGCTGTCCGCGTGGGAGTTACGAGCATTCNTTCAACA |-----|
465 TCTACTTCTGCTCGGTGACGATTCAGATCAACGAGACCGCGCGGACGACATCTGCTGC |-----|
481 486 509 514
525 CGATCGCGCGCCAGCCGGATCCAGCAAGGGAGGTGCGC
598 AACGGAGACCCGNTCCGACCCGGCATCTTCGGCCTGTGCTGTGATGCTGCTCTGA |-----|
648 653
658 TCGCATTCGGCTACAGCGGTTGCCTTCTGGCCACAGGGCAAAACCTACGACGGCTATT
718 TCACCGAGCGCGGTGGATCAACCCCGGTAACCTCGGTTTATGTCTCGGGCTCAAGGTGG |-----|
723 728
778 G

9 matches found in sequence:
ar149055 ; TOIG of: ar149055 check: 8421 from: 1 to: 689
(from "mycobacter.ge.seq")
TOIG of: ar149055 check: 8421 from: 1 to: 689
LOCUS ARI49055 689 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 71 from patent US 6228371.
ACCESSION ARI49055
VERSION ARI49055.1 GI:15113646
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 689)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 71 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..689
/organism="unknown"
BASE COUNT 123 a 224 c 184 g 153 t 5 others
ORIGIN
ARI49055 Length: 689 February 21, 2002 16:59 Type: N Check: 8421 ..
Found using 'seq2-3' (pappu403.key)
1 CTATCCGACAGGCTTCGACAGCGCTCGGCTGNACCGCAGAAATCGGGTGCACCCAGATT |-----|
20 25
61 GCCAGTAGCGCGGC
...
99 TCGTAGCCAAATCATCGGCCCGGTAGTATCTCCGAGATGACAGATGAATGTCGTCGACA |-----|
149 154
159 TTTCGGGTGGCAGTTCCGATCAACGCTCTATCACTTCATTTT
...

213 CTGACCATCGGCGCTGGCCCCNCGTATCGCGGTATGCAAACTGCTGTNGTGCACCG |-----|
263 268
273 ATACCCCGCGCTGTATTCGCCCTACACAAATTCCTTGGCAAAATTGTTCTGATCAACTTGG |-----|
288 293
333 CCATCGGCGCTGGCGACCGGAATCTGCAAGAAATTTCAGTTGGCATGAATCGAGCAGT |-----|
343 348
393 ACTCCGATTCGTCGGCATGCTTCGGCGCCCGCTGGCCATGAGAGGCGCTGGCGGCT |-----|
407 412 419 424
453 TCTTCTTGAATCCACTTCAT
...
486 TCTTGGCTGGAACAGGCTGCCCGGCTGTGTGATCTGCGCTGCATCTGGATCTGCGCA |-----|
536 541
546 TCGCGTCAACGTGTCCGCTTCTTCATCATCGNCGCAACTCCTT
...
603 CGGTGCGCGGCACCTACACACCCGACCCGGGCGTGCAGTTGAGCAGCATCTGTCGC |-----|
653 658
663 CTGCTGACCAACAACACCGCAGAGCG

4 matches found in sequence:
ar149056 ; TOIG of: ar149056 check: 6755 from: 1 to: 274
(from "mycobacter.ge.seq")
TOIG of: ar149056 check: 6755 from: 1 to: 274
LOCUS ARI49056 274 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 72 from patent US 6228371.
ACCESSION ARI49056
VERSION ARI49056.1 GI:15113647
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 274)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 72 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..274
/organism="unknown"
BASE COUNT 50 a 90 c 84 g 50 t
ORIGIN
ARI49056 Length: 274 February 21, 2002 16:59 Type: N Check: 6755 ..
Found using 'seq2-3' (pappu403.key)
1 CCGCAGCAGCAGGCAAGCATCGACCGCGTGGATTCCCGCCATCCCGGCGCATGATGAT |-----|
47 52
61 CATGTCCGACACCGCGCCGACCTGCTTCCGAGTTGACCGGCTGCCTGTCGACGC |-----|
74 79 116
121 CGCAACGATCGTTCGTTCAATGATCCGCGCTCGAAATTGGCCATGGCG
121

```
...
353 ACTGGTGTGGCAATCTGTGAATACACACCGTTGCACGGGATTTTACCGTCCGCGG 403 408
413 ATCGGGCTGTGGAACACGGGGCGGTGATTTGCCCGCTGACCCGCATGCTGACCGCGGTG
473 GCATGTTTTTGGGCTTCTGCTCCCGGGTTGTGCTCTCGCCCGTCACGTACGCGCAC 491 496
533 ATCGAATTGCTGGAGTGTCTGCGCGCTGATCGCAATGTTCATTCACCGCTTGSTATGC 553 558
593 GCGGCTACAGGCGGCGTTGCCACCTCGACCTGACATGACATGACACACNGTCCCGTACC 593 598
653 CTGCGCCTGCTGTGTATACGCTTCGCGACGCTGTCTGGGTTGGAGCGCGTTGCG 673 682 687
713 CCACGCGTTTGGCGG

7 matches found in sequence:
ar149053 : TOIG of: ar149053 check: 1693 from: 1 to: 1028
(from "mycobacter-ge.seq")
TOIG of: ar149053 check: 1693 from: 1 to: 1028

LOCUS AR149053 1028 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 69 from patent US 6228371.
ACCESSION AR149053
VERSION AR149053.1 GI:15113644
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1028)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL peptides
FEATURES Patent: US 6228371-A 69 08-MAY-2001;
Location/Qualifiers
source 1..1028
BASE COUNT 204 a 330 c 325 g 162 t 7 others
ORIGIN
AR149053 Length: 1028 February 21, 2002 16:59 Type: N Check: 1693 ..
Found using 'seg2-3' (pappu403.key)

...
87 GGAGACCGCGGCCAACACAGGTGCCGTACATGACCCGACGCGCATTCATCGCGGAGC 137 142
147 CGGCGTAGATGTTTTCTGCAAGGCGGTGCGGCGTGACCCNTCCGCGCGACACCGGCA 191 196
207 CCTTCCCGCGTCCACGTGCGGCTGGGTGGTGGAGCGCGGACGACACCCACCAATGATGGA 238 243
267 CATGGCTGTGGGTGATGATGACCGSCG

...
327 TACAGTCCAGCGCGCGGCCACCTCTGTGTGACACCAACGGGTGATGAGATCCAC 377 382
387 CCAGTGTACACCTTCANAGAAAGCTGATATTGAGATATGCAATCCG
...
501 AGGCTGGATATGACCGATGTGCGGCGGACCGTGCAGAAACGAGTACGCTGTGTC 551 556
561 CACACACCGCGACCATTCGCGACGCTTGATACACACGCGGACAGCGCGCAATGATCCG
621 CGATGCGCGCTGCTGCAGAAATCCGTTGTGTCATGCAACGAGTAAAGAGTGTACACG 626 631
681 GCCT 629 634

12 matches found in sequence:
ar149054 : TOIG of: ar149054 check: 7262 from: 1 to: 780
(from "mycobacter-ge.seq")
TOIG of: ar149054 check: 7262 from: 1 to: 780

LOCUS AR149054 780 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 70 from patent US 6228371.
ACCESSION AR149054
VERSION AR149054.1 GI:15113645
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 780)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL peptides
FEATURES Patent: US 6228371-A 70 08-MAY-2001;
Location/Qualifiers
source 1..780
BASE COUNT 151 a 245 c 244 g 134 t 6 others
ORIGIN
AR149054 Length: 780 February 21, 2002 16:59 Type: N Check: 7262 ..
Found using 'seg2-3' (pappu403.key)

...
45 NAGGTGATCCGTTGGCAACGTCTGCCCGACCGGTGCTTCGCGCGCGCTGGGC 95 100
105 GCACGCGACACGCTGATGCGGAGGTAATCACCAACTCAACGCGGTGCTGGCGCGTC 155 160
165 GATGCAAGAGCGCGCAATTNTGCGGCAAGTGTGACCAAGCTGACGAGCTGTGACGGG
225 CCTGGCCAAGAACCGGATCCGATCGCGGCGCAATTTTCGCGCTGTGCGTGCACGACGAC 253 258
285 GGATCTTACGGAAGTGTTCGGGAATTCGCGCGCGCGCTGCAAGGATCTGGAACGCGC 340
```

```
206 ACAATTAGCTGACTGATTGAGACGCTGCCGAGTCGCTCAACACACGCTCGACGAGCT
      227 232 239 244
266 GCACGAGAACTGGAAAGTAGTTTCA
-----
16 matches found in sequence:
ar149051 : TOIG of: ar149051 check: 2309 from: 1 to: 1306
(from "mycobacter-ge.seq")
TOIG of: ar149051 check: 2309 from: 1 to: 1306

LOCUS AR149051 1306 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 67 from patent US 6228371.
ACCESSION AR149051
VERSION AR149051.1 GI:15113642
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1306)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 67 08-MAY-2001;
FEATURES
source 1. 1306 Location/Qualifiers
BASE COUNT 251 a 376 c 407 g 271 t 1 others
ORIGIN

AR149051 Length: 1306 February 21, 2002 16:59 Type: N Check: 2309 ..
Found using 'seq2-3' (pappu403.key)

1 GTGATACAGAGGCGCCCAACAGTAGACACCTCGCGGGCCAGCTGTTGCAACAGCTTGTCG
      12 17 41 46 50 55
61 CAGTCGAGGCTCAGCGCGGCTCCGAGAGGCTGCCCGTGTTCGACAGCTGGCGCTC
      114 119
121 GCAGCAATGCTGGGGCATTGGCCGTCACGTCAGTGATGACAGCTGGT
      114 119
...
200 GCATTACCCCGGAGGACACCTCAATCGAGAAGTGTGATCGGGGGTGAATCGCTCC
      250 255
260 TGGCGGTGGGGTATCGTGTGGGCTCATCTTCTGTGTCGGCGCT
      693 698
...
643 CGGGGCTCAACACCGAGGACCGGACCTACCTGAATTTGACAAGGTGAGACGCTGGGC
      693 698
703 ACCAGCACCGAAATTCGGTGTGTGCTGCCGTCGGCAAGCGTA
      853 858
...
803 TCTTGTCAACGCTGACGTGATGCCCTAACCCGGTGGCAACAACACTCGTCAACGCTTCC
      853 858
863 AGATCGAAGAATCACCAAGACCGAGCATTCGTGGGCCACTGGCC
      853 858
...
917 GTGCGACGTATCACTGATGATGAACCTTCGAGGTCCGGCTGTGACCCCAACGATTTCA
      853 858
```

```
967 972
977 AGCCCTACCTGCAGCAGACGCATCGACGGGAATACAAAGCGCGAGGCGCTCGGGGATCA
      1012 1017 1029 1034
1037 ACCAGCCGCCCTTCGCGGTGACCAACCAACCGTTGATTAATCTGCGCGGTGAATTGGCCC
      1114 1119
1097 CGCAGCCCGTAGGTTAGACGCTCATGCAATATGAAGCCGACGCTTTGAGTTGTGCGC
      1151 1156
1157 GCGTTCTTCGTGTGTCAGCGCGGTCTGTACGCGCTGTGACCTCGATGTGCGCACCGGT
      1246 1251 1264 1269 1276
1217 GGTTCGAGTGGGCTGCGACCACTCGCTGCGCTTACCGCGCGCATGCGCTGATTCGTC
      1246 1251 1264 1269 1276
--
1277 GCCACCTTCTTCGCGTTGTGCGCGGAT
      1279
-----
12 matches found in sequence:
ar149052 : TOIG of: ar149052 check: 9681 from: 1 to: 728
(from "mycobacter-ge.seq")
TOIG of: ar149052 check: 9681 from: 1 to: 728

LOCUS AR149052 728 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 68 from patent US 6228371.
ACCESSION AR149052
VERSION AR149052.1 GI:15113643
KEYWORDS
KEYWORDS AR149052.1 GI:15113643
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 728)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 68 08-MAY-2001;
FEATURES
source 1. 728 Location/Qualifiers
BASE COUNT 101 a 243 c 225 g 149 t 10 others
ORIGIN

AR149052 Length: 728 February 21, 2002 16:59 Type: N Check: 9681 ..
Found using 'seq2-3' (pappu403.key)

...
39 TGGTCTGTAGAGGATCCGCGCGGATGGCCAGTCTGTGCGGGGTTGTGCGGCACAGA
      89 94
99 TTGCGCGCGCGGCTGAAMCCCGAGAGCGCGCATGCGCTTGCCACACGAGATCGGCTGAC
      120 125 123 128
159 ACCGAGCGAACCCTGACAGATCTTGTGACGAGCCTGTGGGCMCTTGTGCGGCGCTCG
      218 223
219 GCGACCATCGGTGCGCATCGAAGCGCGCTGCAACCCAGGCAATCCACAGSTGCA
      223
```

```
331      CCGCAGGATGATGACGGAGCTTCTTCCCTTCTACTGTGTGGAGCTGCTGCTCAAGATC
      346 351
391      AACGGCAAGGGCGCGCAGCCGGTGTACATTCACATGCGC
      373 378

4 matches found in sequence:
ARI49048 ; TOIG of: ari49048 check: 6186 from: 1 to: 481
(from "mycobacter-ge.seq")
TOIG of: ari49048 check: 6186 from: 1 to: 481

LOCUS      ARI49048      481 bp      DNA
DEFINITION Sequence 64 from patent US 6228371.
ACCESSION  ARI49048
VERSION     ARI49048.1 GI:15113639
KEYWORDS
SOURCE
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 481)
AUTHORS      Nano, F.E.
TITLE        Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
             peptides
JOURNAL      Patent: US 6228371-A 64 08-MAY-2001;
FEATURES
             Location/Qualifiers
             source          1..481
             BASE COUNT     98 a 116 c 146 g 113 t 8 others
ORIGIN
ARI49048 Length: 481 February 21, 2002 16:59 Type: N Check: 6186
Found using 'seq2-3' (pappu403.key)

1      KGTCTCGCGNCTTATACATCCGGTGCCTCCANCGTAATCTGCGTGTGATGCCGTCGCGG
      23 28
61      AANTATAGCAAAATGGCC
      337 342
...
133      GGGGCGTGAGGTGGAATAATGCGGTGTGCGGTGATATTTCGACGCGCAAGCCACATA
      183 188
193      TAGTGTGATCGACGCGCAATAAASACACGCTCTGGCCACGTTTCTTG
      439 444
...
287      TTGCAGATGCGCGCAAGCGCATCAGCTAATGACGATTCGGGCGCCACAACGTTCC
      337 342
347      GTTCTGGCGGTTTCTCTGACTACACTCAGATCAAGGTGTGATTTTGATGATCGCC
      439 444
407      AGTTATCGGGGCCAATGTCGACGATGATGACGTGAGACCAATGTTCTTTTGGCAATN
      439 444
467      CGGGTGCGAATAGCG

3 matches found in sequence:
ARI49049 ; TOIG of: ari49049 check: 1123 from: 1 to: 469
(from "mycobacter-ge.seq")
TOIG of: ari49049 check: 1123 from: 1 to: 469

LOCUS      ARI49049      469 bp      DNA
DEFINITION Sequence 65 from patent US 6228371.
PATENT      08-AUG-2001
```

```
ACCESSION  ARI49049
VERSION     ARI49049.1 GI:15113640
KEYWORDS
SOURCE
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 469)
AUTHORS      Nano, F.E.
TITLE        Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
             peptides
JOURNAL      Patent: US 6228371-A 65 08-MAY-2001;
FEATURES
             Location/Qualifiers
             source          1..469
             BASE COUNT     96 a 130 c 141 g 97 t 5 others
ORIGIN
ARI49049 Length: 469 February 21, 2002 16:59 Type: N Check: 1123
Found using 'seq2-3' (pappu403.key)

1      GCGGAGTCACTGTAAGCCGAGGAGCGGAAGAGCGCCCATACGAAACCGCTCTCCC
      34 39
61      CCGCGCTTGGCCGATTCATTAATGCAGC
      320 325
...
270      ATACGACTCATTATAGGCAATTCGAGCTCGGTACCCGGGGATCTCTAGAGTCTGCTGCG
      320 325
330      TTGGCGGCGACGACGATGATCCACGGTGTGCGCGCGCGGCGGCDPTATACACGCGCGG
      335 340
390      G
      7167 7167
...

3 matches found in sequence:
ARI49050 ; TOIG of: ari49050 check: 7167 from: 1 to: 291
(from "mycobacter-ge.seq")
TOIG of: ari49050 check: 7167 from: 1 to: 291

LOCUS      ARI49050      291 bp      DNA
DEFINITION Sequence 66 from patent US 6228371.
ACCESSION  ARI49050
VERSION     ARI49050.1 GI:15113641
KEYWORDS
SOURCE
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 291)
AUTHORS      Nano, F.E.
TITLE        Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
             peptides
JOURNAL      Patent: US 6228371-A 66 08-MAY-2001;
FEATURES
             Location/Qualifiers
             source          1..291
             BASE COUNT     77 a 76 c 67 g 50 t 21 others
ORIGIN
ARI49050 Length: 291 February 21, 2002 16:59 Type: N Check: 7167
Found using 'seq2-3' (pappu403.key)

146      CAAGCTGATATAGAGTCCGCGTGTGACTATTGTATGCCCGCCAGCAGGATTGCGG
      196 201
```

2185 2190 2227 2232
2230 2235
2244 CCTGCCCGGATGCGGGTGTMCGGTGTTCGAGGGTTGCCGGCCGATCGGTACCC
2304 ATACATCGTCACCGCGCGGAGCGCATCTGATCACCGAGCTCGG
2308 2313

12 matches found in sequence:
ari49046 : TOIG of: ari49046 check: 4144 from: 1 to: 821
(from "mycobacter-ge.seq")
TOIG of: ari49046 check: 4144 from: 1 to: 821

LOCUS ARI49046 821 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 62 from patent US 6228371.
ACCESSION ARI49046
VERSION ARI49046.1 GI:15113637
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 821)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 62-08-MAY-2001;
FEATURES
source 1..821
location/Qualifiers
BASE COUNT 132 a 292 c 267 g 125 t 5 others
ORIGIN

ARI49046 Length: 821 February 21, 2002 16:59 Type: N Check: 4144 ..
Found using 'seq2-3' (pappu403.key)

...

26 AGGCCAACCTGACCGCGCTCAGACACGGGCTGTCCAGCTGTGCGATGCGCCGACGACGTGC
76 81
86 TGGCCGAGCCCGACCCCAATGCGGCGATGTCAACCGGTTCCGGGCCAGGCGTGGCGAGC
135 140

146 CGGACGGACGCTGGCGGTATCAGTCCCGTCGGCTTCAMACCCGAGGCGTGGCGAGGA
152 157

206 CC

208 TCAAGTCCGANCCTGGTGTCTCCAAACCCGGCGTGTCAACTCCGATGCTCCGCCAAC
258 263

268 AAACCCAAAGCCCGCATCACCGACTCCGGGGCAACCGCCGAGGAGGAGGCCCGGANTCGG
274 279

328 GATTCACAGGGTTGCNACGGCGCTGCCGTTCCNGATTGAGACCCGGACGTACCCCGGT
348 353

388 GATGGCAGCTACGGG

...

444 AGTTACGCCCCCGAGCCCGGCGCGCTGTGTGTTCGCCGGCGGCCCATCT
494 499

504 GGTCTACAGAGAGACGCGCATTTTCATCTACGGCCAGTCCCTGAACCTGCAGTGGGCG
521 526
564 TCACCGGCGCGGAGCGCCGATCCAGCCACTGGGGCAGGTATTTCGATCGACATCGGAC
565

624 CGCAACCGCGGTGGCGCAATCTGGGTTTCGGCTGGCCCTGGGCGCGCGGAGGCGGACG
664 669

684 TGGCGCGCATTTGTGCCCTATGACCGCAACCTGAGCCCTGAGCAATGTTCCCTTCACCC
695 700

744 CGCCCGGGTTCCGGTCTGGAATCTTCAGCGGTTGATCGGTCACGACCGCGTGT
790 795

804 TGATGACATCGCACCG

8 matches found in sequence:
ari49047 : TOIG of: ari49047 check: 7004 from: 1 to: 479
(from "mycobacter-ge.seq")
TOIG of: ari49047 check: 7004 from: 1 to: 479

LOCUS ARI49047 479 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 63 from patent US 6228371.
ACCESSION ARI49047
VERSION ARI49047.1 GI:15113638
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 479)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 63-08-MAY-2001;
FEATURES
source 1..479
location/Qualifiers
BASE COUNT 85 a 158 c 145 g 78 t 13 others
ORIGIN

ARI49047 Length: 479 February 21, 2002 16:59 Type: N Check: 7004 ..
Found using 'seq2-3' (pappu403.key)

...

47 TGTGTGGGCGNTGCTGAGNCCTACACNGATGCTTGGACGAGCGGTGACGCTGCTA
97 102
107 TCAGCGTTGATTTCACCGCGCTGCGCAACGACGACATCTCCAAAGCGCGTGGCTAC
109 114

167 ACCAAGCGCGCGCGCTCGGTGCGCATCTNCTGTGCGAGGCTCGCGCANNCGTTGGCGA
187 192

227 ANGTGTTCCGAGAC

...

271 GACCACGACTACCTGACAACTGTGCTGACAGCGCTGCGGAGCAATAACAGGCGCTGTGTC
321 326

9 ATTGGAGCTTGGAGACCGCGTTGGCGGTGGTGTCTGATCATGAACTTACGACGCGCCACC
59 64
69 GCAGCTTCCCGTCAATGGGTGCTATATCCCTGTGGCGCTGCCATGATGTTCTCGAAGT
127
129 ---|
132 GATTACGCGTCTGGCAGCGAGTACACCGAGGGTATGCCGCCAACCATCG
...
186 TGTCCGGGTCACTCACGGCTGACCGTGTTCGGCTGTGCGCGCACCATTCGTGGTG
236 241
246 GCGGATTGGCGCGGAGTCCGAATTCTGTGACCCACCTGTTCAGCTGCCGGCGCT
306 |-----| |-----| |-----|
TGTTCGTGCTGTGCGCATCACATCGCTTTCGCTGTGATGCGCATTCGCGCT
311 316 329 334
314 319 335 340
366 GGTTCGAGGTGACCAAGCGGTGAGTCCCGCCACATTGAGCTACCCGCGGATGAGGCA
426 |-----|
GACTACGGCGACNGCTGGCGGAGGAACTCAAGAACTCGGGGAAACACTCCGACACCG
432 437
486 TT
...
513 TGGGTAACCTGCACCATCAAGTATGTGCGCTTTCGTCTGTATCCGCGTTTGT
563 568
573 --|
575 GCCAAGGCGCAGCAAGCAACGCGGTGCAATTGGGCATGTGGCGCTGAT
...
720 GTGCTAGTTACCGTGTATGCAATCGCGCGCGGTGGCGCGACGCTGGCGACAGCG
770 775
776
-|
780 APTGCAACCCCTGATCAGCGGCAAGGTCAGTGCATTTCTAAAGCCTCCGTGACCGCTCG
831 836
781
840 TTGACAGACAGACTGCGCGAGAGTGGCGGCATCGGGTTTGGCGTTCCGAGTCACT
883 888
900 CTTACGCTGGCTGGGTCTGGCGCGCGGTGGCGCTG
...
961 TGGGCTTCACTGCGGTGAGCGCGCTGCTGATCTGGGTCTGGCTCAGACCATTCAGCT
1011 1016
1021 TCCGCGCGGATTCGCTGATCCCTGGCGTGGCGGTATCGCGCGGTGATGGCGGCAAG
1026 1031

1081 AAACCAACCCGTCGTGTGGCGGCGGTGAGCGCGGTGTGCGCAACCTGCG
1106 1111
1141 GGTGATCCGTGATTTCTGCTCTCGGTGGCGCGCGGTGATGCTGCTAGTACG
1201 CGGACACGGTCCGCAAGCAACCCGAGATCAGCGCTTACTCGCACGGGCACTGACCCGGT
1229 1234
1261 GGGGCCCTATTTTGTAAGTGTGACCTGTGACCTGTGACGACTGTGACACCCGACGCA
1295 1300
1321 GGGCGAATTGCCGGTAAAGGCAAGCTATCCCGGTGACGCTGTGATACCGGAAGTATTC
1341 1346
1381 |-----| |-----|
CCGGCGCGGTGCGCTTGTGCTGACAGTATACAGAGACCCCGCAACACCGACGACCTT
1384 1389 1397
1441 GTTTCGG
...
1478 ACTGTGACCCCGCAGCGCGGGCGCTGACCGGATTTGTGCTGATGCTGTGAGCTTGGTG
1528 1533
1538 GTGACCACTGGGTGAACCTACGCGAGCAGNTCCGACCGGAAATGCTCGGTGCTTATC
1598 TTTGACGAGGCGCGGCTCGACGGGAGCTTAAGCGGCTGCGGCAACCGTCCGAG
1623 1628 1640 1645
1658 AGCCGCCAGACCCGGGCAATCGCGCATCATGAA
...
1740 CAGCGCGGCTGGAGGCTGCGGGCCAGCGCGCGCAGACTGCGCGGTGAGCGCGCTG
1790 1795
1796
-|
1800 CTGGCGGTGTTGTTGATCGCGTGCACACACCGACCGGCTGCCGACGCTCTGAGCGCG
1801
1860 GTGCTGAGATGATGATCCGCCCGCGCTTGTGCTTGCAGCGCCAGCGCTTGGC
1899 1904 1911 1916
1914 1919
1920 GCGCGCAAGATGAGAGAGAGCGCGCGGTGTGTGTGTGCTGCGCTTGGC
1964 1969
1980 GCCGAGACATTCACAGCTGAGCTTGTGCTGACCGGCTCGG
...
2124 AACACGACTGCGCGGACCGCGCTGTGGCGGAGTGTGAGCGCGCGCGCGCG
2174 2179
2184 GGGCGACACCGGCTGTTCCGGATCAGACGCTGCTCACCTCGCGCTGTGCGCGTGGC

```

1 match found in sequence:
ar149042 ; TOIG of: ar149042 check: 2098 from: 1 to: 262
(from "mycobacter_ge.seq")
TOIG of: ar149042 check: 2098 from: 1 to: 262

LOCUS ARI49042 262 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 58 from patent US 6228371.
ACCESSION ARI49042
VERSION ARI49042.1 GI:15113633
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 60 a 71 c 73 g 45 t 13 others
ORIGIN

ARI49042 Length: 262 February 21, 2002 16:59 Type: N Check: 2098
Found using 'seq2-3' (pappu403.key)

...

131 CGGTGCTCAGAGCCCTGATGAGTCTTCCCGGACGTAAGCCTAAGACGACATGG |-----|
181 186

191 CGCGGAGCTTGCCGAGCANCTGCGGTATCCCGAGACCTGTTCATA
-----
1 match found in sequence:
ar149043 ; TOIG of: ar149043 check: 6253 from: 1 to: 241
(from "mycobacter_ge.seq")
TOIG of: ar149043 check: 6253 from: 1 to: 241

LOCUS ARI49043 241 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 59 from patent US 6228371.
ACCESSION ARI49043
VERSION ARI49043.1 GI:15113634
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 45 a 92 c 54 g 40 t 10 others
ORIGIN

ARI49043 Length: 241 February 21, 2002 16:59 Type: N Check: 6253
Found using 'seq2-3' (pappu403.key)

...

149 TCGACAGCTGTGCTGCGCCGCGACACGACCTGAGACATCTACCCCATCGTTGAC |-----|
199 204

209 GCGAGCTGGCAGCTCAGACCGGTTTCGGTGCCG

```

```

3 matches found in sequence:
ar149044 ; TOIG of: ar149044 check: 1735 from: 1 to: 243
(from "mycobacter_ge.seq")
TOIG of: ar149044 check: 1735 from: 1 to: 243

LOCUS ARI49044 243 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 60 from patent US 6228371.
ACCESSION ARI49044
VERSION ARI49044.1 GI:15113635
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 30 a 83 c 78 g 37 t 15 others
ORIGIN

ARI49044 Length: 243 February 21, 2002 16:59 Type: N Check: 1735
Found using 'seq2-3' (pappu403.key)

...

57 NTNCCGCGGNCCAGNCCAGNCTNGATGTCNGCTATATACTGCGGATCGCGCGG |-----|
107 112

117 GCGTCCCGGACACACAGGTGNGCGCCGCCGCTGCTTCCGCCCAATTCTGGGTGNCGACATN
-----
1 match found in sequence:
ar149045 ; TOIG of: ar149045 check: 3530 from: 1 to: 2348
(from "mycobacter_ge.seq")
TOIG of: ar149045 check: 3530 from: 1 to: 2348

LOCUS ARI49045 2348 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 61 from patent US 6228371.
ACCESSION ARI49045
VERSION ARI49045.1 GI:15113636
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT 335 a 760 c 811 g 433 t 9 others
ORIGIN

ARI49045 Length: 2348 February 21, 2002 16:59 Type: N Check: 3530
Found using 'seq2-3' (pappu403.key)

...

```



```
80 TCCTACACACGGGGGGGGCGTTGGCAGGGCCGACACCCCACTACCTGNCGGGG
138
140 ---|
CGCCACCGCTTGGCCGTTGCTGAGACCGGACTTCCCGGACCGCTGATGTCACGCGTC
143 |--
197
200 ---|
GCCGCTTAATCCCGATGGGCGCCCGCCGACCGCGGCATCTAAGTCTGGGCGCGGG
202 |----|
227 232
260 CGAGCGGCTCCGCTGTTCCGG
...
384 TCTTCGACATCCGSAGCTGCGACTGTCGAAACTGTCTGCAAGAGTAGTGTGTCGGG
434 439 |----|
444 GGTGTGTGTGCTTTGGCGGCTGTCGGCGCTGCGCGCGCGCGGCTTACCGGAAAC
504 TGACTACCACTACCGTGTGCTGCTATTTNCTSTAGGCGCTGCGCGCTGACCCAGAGAC
539 544 |----|
564 AAGTCAGATCANGGGTGTGGGGTCTGCTTATCAGACAGATCAGCGCGCGCGAC
618 623 |----|
624 AAGATCGAGTCAGCTTGCACTACAGCAACAATACAGTGTGCGCGCGAC
...
1 match found in sequence:
ar149039 : TOIG of: ar149039 check: 4691 from: 1 to: 117
(from "mycobacter-ge.seq")
TOIG of: ar149039 check: 4691 from: 1 to: 117
LOCUS AR149039 117 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 55 from patent US 6228371.
ACCESSION AR149039
VERSION AR149039.1 GI:15113630
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 117)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 55 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..117
BASE COUNT 20 a 47 c 31 g 19 t
ORIGIN
AR149039 Length: 117 February 21, 2002 16:59 Type: N Check: 4691
Found using 'seq2-3' (pappu403.key)
...
9 TCGTTCGCCGCGACATGACTATCAGCCGACCGGCACTGCTGACCTGATCGGCACAC
59 64 |----|
69 GCTGAGGCGCTACCGGCTGCAATTCAATCAACCCGCTGCGGGTGC
...
1 match found in sequence:
```

```
ar149040 : TOIG of: ar149040 check: 2574 from: 1 to: 242
(from "mycobacter-ge.seq")
TOIG of: ar149040 check: 2574 from: 1 to: 242
LOCUS AR149040 242 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 56 from patent US 6228371.
ACCESSION AR149040
VERSION AR149040.1 GI:15113631
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 242)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 56 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..242
BASE COUNT 47 a 87 c 63 g 45 t
ORIGIN
AR149040 Length: 242 February 21, 2002 16:59 Type: N Check: 2574
Found using 'seq2-3' (pappu403.key)
...
1 AGGTGCTGCTTCAATGCTGCGCCCAATCAGTTTCTACACCGACTGTACACCTT
21 26 |----|
61 CGCAGACAACGCGCCA
...
1 match found in sequence:
ar149041 : TOIG of: ar149041 check: 3152 from: 1 to: 345
(from "mycobacter-ge.seq")
TOIG of: ar149041 check: 3152 from: 1 to: 345
LOCUS AR149041 345 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 57 from patent US 6228371.
ACCESSION AR149041
VERSION AR149041.1 GI:15113632
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 345)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 57 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..345
BASE COUNT 83 a 90 c 105 g 67 t
ORIGIN
AR149041 Length: 345 February 21, 2002 16:59 Type: N Check: 3152
Found using 'seq2-3' (pappu403.key)
...
198 CGCGGCTCAACACCGAGACCGGACCTAAGTATTCGACACAGCTGAGAGCTTGGGC
248 253 |----|
258 ACCAGCACGAATTCGGTGTGCTGCTGCGCTCGGCAACGCTA
...
1 match found in sequence:
```

```

269 TTGTTCCCGTGATCGGCGAGCTGGTGNCCCCGGGCCCTGTTGGCACTGCGCAGACG |-----|
329 TGC 319 324 322 327

3 matches found in sequence:
ar149036 ; TOIG of: ar149036 check: 1340 from: 1 to: 507
(from "mycobacter_ge.seq")
TOIG of: ar149036 check: 1340 from: 1 to: 507

LOCUS AR149036 507 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 52 from patent US 6228371.
ACCESSION AR149036
VERSION AR149036.1 GI:15113627
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 507)
TITLE Nano, F.E.
Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 52 08-MAY-2001;
FEATURES
SOURCE 1..507
Location/Qualifiers
/organism="unknown"
BASE COUNT 94 a 150 c 140 g 123 t
ORIGIN

AR149036 Length: 507 February 21, 2002 16:59 Type: N Check: 1340 ..
Found using 'seq2-3' (pappu403.key)

1 TGTATTCCCGTGATCGGCTTGCTGCAGGTAGGSCCTTGAATCGTTGGGGGTACGACGCGG |-----|
61 GACCTCGAAGTTTCATCATCGAGTATACGTGCCAC 40 45

104 GCGCAGTGGCCCGACGATGCTCCGGTCTTGGTATTTCTTGATCTGGAAGACGTTGACC |-----|
154 159

164 GAGTTGTTGGCCACCGGGTTAGCATTCACGTCACGCTTGAACACAGA

264 TACGCTTCCCGGACGGCAGCACACACGGAATTTCCGTCTGTGTCGCCCAACGCTTGA |-----|
314 319

324 CCTTGTGGAATTCAGTAGGTCCGGTCTCTGGTGTGAGCCCGCGG

4 matches found in sequence:
ar149037 ; TOIG of: ar149037 check: 5460 from: 1 to: 293
(from "mycobacter_ge.seq")
TOIG of: ar149037 check: 5460 from: 1 to: 293

LOCUS AR149037 293 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 53 from patent US 6228371.
ACCESSION AR149037
VERSION AR149037.1 GI:15113628
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 293)
TITLE Nano, F.E.
Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 53 08-MAY-2001;
FEATURES
SOURCE 1..293
Location/Qualifiers
/organism="unknown"
BASE COUNT 68 a 104 c 73 g 48 t
ORIGIN

AR149037 Length: 293 February 21, 2002 16:59 Type: N Check: 5460 ..
Found using 'seq2-3' (pappu403.key)

1 CCACACAACACAAATCTACGTGTATGTCAGTCGTAGTCCATCCGACGTGATGCAAG |-----|
61 GACACACCCGACGGCCCAACGCAATATACATGCTGCGCTCCGGCTCACAAACACATCAT 90 95 46 51

121 CATGGACTGTTCCTACTACGGCGTAC

174 CGATGTGCTAGTGGCCACCCAGATCTCTTCTACACGCGTGTCTTCTGCTACGACGCGGTGG |-----|
224 229

234 TCGGTGGGGGCTCAGAGGCCACACACACACACACCATGCGTGCCTGAAGCTACGCCGACG |-----|
279 284

294

7 matches found in sequence:
ar149038 ; TOIG of: ar149038 check: 3192 from: 1 to: 820
(from "mycobacter_ge.seq")
TOIG of: ar149038 check: 3192 from: 1 to: 820

LOCUS AR149038 820 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 54 from patent US 6228371.
ACCESSION AR149038
VERSION AR149038.1 GI:15113629
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 820)
TITLE Nano, F.E.
Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 54 08-MAY-2001;
FEATURES
SOURCE 1..820
Location/Qualifiers
/organism="unknown"
BASE COUNT 130 a 284 c 257 g 126 t 23 others
ORIGIN

AR149038 Length: 820 February 21, 2002 16:59 Type: N Check: 3192 ..
Found using 'seq2-3' (pappu403.key)

20 CGGTGCGGAGGAGGTACACCAAGCANTCCGGACCGCGCTTCYCGCGGGAACGCGTGC |-----|
70 75

```

```
411 TGTGATCTTGTCCA
...
5 matches found in sequence:
ar149033 ; TOIG of: ar149033 check: 426 from: 1 to: 423
(from "mycobacter-ge.seq")
TOIG of: ar149033 check: 426 from: 1 to: 423

LOCUS ARI49033 423 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 49 from patent US 6228371.
ACCESSION ARI49033
VERSION ARI49033.1 GI:15113624
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 423)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 49 08-MAY-2001;
FEATURES
source 1..423
Location/Qualifiers
BASE COUNT 89 a 139 c 109 g 77 t 9 others
ORIGIN
AR149033 Length: 423 February 21, 2002 16:59 Type: N Check: 426 ..
Found using 'seq2-3' (pappu403.key)
...
33 TACCACGTGACATCGACGACGAGCGGCTGCANTCAGCNCNCTTCGTCACACCGCGCATCTC
83 88
93 AACNTGCACCTAAGCGCACTGTGNCAGCACCCCGGACCATGCGCATCTG
...
225 AATGGGTTKCCCGTTGTCACGAGAGATCCCGCTGATNCATGGGCAATACGCTCCCT
275 280
285 ATCCGACGTGTACACAGTGGGTGCAAGCGGCGGATGGCCAAAGTTGGTGGTCGACCAACCCCG
313 318 329 334
345 ACTAGCTTTTCACAACACTCGACCTGACCGGTGGAACATCAACCCGCGCATGTGATGCCAG
389 394
405 CAACCTATGTGGGATCTG
-----
6 matches found in sequence:
ar149034 ; TOIG of: ar149034 check: 5923 from: 1 to: 279
(from "mycobacter-ge.seq")
TOIG of: ar149034 check: 5923 from: 1 to: 279

LOCUS ARI49034 279 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 50 from patent US 6228371.
ACCESSION ARI49034
VERSION ARI49034.1 GI:15113625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 279)
AUTHORS Nano,F.E.
```

```

TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 50 08-MAY-2001;
FEATURES
source 1..279
Location/Qualifiers
BASE COUNT 52 a 92 c 82 g 53 t
ORIGIN
AR149034 Length: 279 February 21, 2002 16:59 Type: N Check: 5923 ..
Found using 'seq2-3' (pappu403.key)
-----
4 matches found in sequence:
ar149035 ; TOIG of: ar149035 check: 4456 from: 1 to: 331
(from "mycobacter-ge.seq")
TOIG of: ar149035 check: 4456 from: 1 to: 331

LOCUS ARI49035 331 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 51 from patent US 6228371.
ACCESSION ARI49035
VERSION ARI49035.1 GI:15113626
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 331)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 51 08-MAY-2001;
FEATURES
source 1..331
Location/Qualifiers
BASE COUNT 30 a 98 c 124 g 64 t 15 others
ORIGIN
AR149035 Length: 331 February 21, 2002 16:59 Type: N Check: 4456 ..
Found using 'seq2-3' (pappu403.key)
-----
1 CGGCGCCGCGCCGCTGTAACCTTGGMGMTGGGTGAGCCAGCTGCCACACAC
10 15
61 GGRAC
...
158 CGGCGGTGTTGSGCTNCGATTGCTGGGCTGTGACCCAGCCGCGACGACACCG
108 213
218 GTGGGTGGGCTGGGCTAGCGGCTGCTATTGGGCTGTGTTCTACG
...

```

5 matches found in sequence:
ar149030 : TOIG of: ar149030 check: 7023 from: 1 to: 186
(from "mycobacter_9e.seq")
TOIG of: ar149030 check: 7023 from: 1 to: 186

LOCUS AR149030 186 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 46 from patent US 6228371.
ACCESSION AR149030
VERSION AR149030.1 GI:15113621
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 186)
TITLE
Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 46 08-MAY-2001;
FEATURES
Location/Qualifiers
1..186
SOURCE
/organism="unknown"
BASE COUNT 36 a 53 c 65 g 25 t 7 others
ORIGIN

AR149030 Length: 186 February 21, 2002 16:59 Type: N Check: 7023 ..
Found using 'seq2-3' (pappu403.key)

....

12 TGCCGCCCGTTGAACGTGAGCNGCTGNCNATNGCCCGAGAGAGAACGTCGCCG 62 67
|-----|
|-----|
|-----|

72 AGGAGTATGACACTGGAGAGACCCGGAAGACTATGACGACTATGACGACTATGAGCGCG 92 97
|-----|
|-----|
|-----|

132 CAGACGAGAGAGCCGCGACGTCGTCATCTGCGACGCGCGTTCGCGGTTCGCTT 162 167
|-----|
|-----|
|-----|

7 matches found in sequence:
ar149031 : TOIG of: ar149031 check: 3749 from: 1 to: 409
(from "mycobacter_9e.seq")
TOIG of: ar149031 check: 3749 from: 1 to: 409

LOCUS AR149031 409 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 47 from patent US 6228371.
ACCESSION AR149031
VERSION AR149031.1 GI:15113622
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 409)
TITLE
Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 47 08-MAY-2001;
FEATURES
Location/Qualifiers
1..409
SOURCE
/organism="unknown"
BASE COUNT 84 a 130 c 113 g 72 t 10 others
ORIGIN

AR149031 Length: 409 February 21, 2002 16:59 Type: N Check: 3749 ..
Found using 'seq2-3' (pappu403.key)

1 |-----|
GTGCTGAAGTGTGTGCGAGACCGTNGATCAGACCTATCCGACCTGAGCGCGCCT

1 6 50 55
CGACGGNTGCTAAGTTCCTCCGACACCATCGGCAAGCGCAGCA

61
....

142 TGCCAGCATCTGTGGTGTATGTCAGTAGAGAGTCGACCGCTATTGGTCAACGCTAAGA 192 197
|-----|
|-----|
|-----|

202 CCTGATGCGCGGTTTCACAGAGCGCGCGCGGTCGACGCCCTGCTGGGGAACATCT 207 212
|-----|
|-----|
|-----|

262 CCGCTTTCGNCGCCAGGTGCAAAACCTTNATCAACGCAACCCGAACTGACCATGTG 295 300
|-----|
|-----|
|-----|

322 CTCGNNNAGCTGCCATCCTCANGACCTGTTGTGACCGCAAGAGATTGGCTGAA 295 300
|-----|
|-----|
|-----|

382 ACCCTGACATCTGGGACATTCAGCG 387 392
|-----|
|-----|
|-----|

4 matches found in sequence:
ar149032 : TOIG of: ar149032 check: 7885 from: 1 to: 464
(from "mycobacter_9e.seq")
TOIG of: ar149032 check: 7885 from: 1 to: 464

LOCUS AR149032 464 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 48 from patent US 6228371.
ACCESSION AR149032
VERSION AR149032.1 GI:15113623
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 464)
TITLE
Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 48 08-MAY-2001;
FEATURES
Location/Qualifiers
1..464
SOURCE
/organism="unknown"
BASE COUNT 59 a 130 c 162 g 104 t 9 others
ORIGIN

AR149032 Length: 464 February 21, 2002 16:59 Type: N Check: 7885 ..
Found using 'seq2-3' (pappu403.key)

....

85 AATGCTTGCGCGGTGTGTTGATGATGCTGCGACACCATGCGACGATCCCGCCN 135 140
|-----|
|-----|
|-----|

145 TGCGCGGATCGCGGCTTNGCTGCGCTGCGCCCTGTGTGGTGT 135 140
|-----|
|-----|
|-----|

231 GGCATCAACNTTCTGGCTGTGCTGTGATTTGGTAACCGTGTGACGACGATG 281 286
|-----|
|-----|
|-----|

291 CGCCCGGCTGCGACTGCGGCTTGTCTTCCTGTGCGCGCGCGGATCTCGGCGCGGTG 284 289
|-----|
|-----|
|-----|

351 TGACTGGGTACCTGCGCTGAGACGACCGGACCGGCACTTCAATCGGATTCGCGGAG 370 375
|-----|
|-----|
|-----|

AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 42 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..221
BASE COUNT 47 a 73 c 37 g 40 t 24 others
ORIGIN

ARI49026 Length: 221 February 21, 2002 16:59 Type: N Check: 5999 ..
Found using 'seq2-3' (pappu403.key)

158 CTCTKCGCCGGCTGCAACACGCATCGTGGCCGCGKACCCGGATGTTTACACCGGC |-----|
208 213

218 TGCA

2 matches found in sequence:
ari49027 ; TOIG of: ari49027 check: 1669 from: 1 to: 117
(from "mycobacter-ge.seq")
TOIG of: ari49027 check: 1669 from: 1 to: 117

LOCUS ARI49027 117 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 43 from patent US 6228371.
ACCESSION ARI49027
VERSION ARI49027.1 GI:15113618
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 117)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 43 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..117
BASE COUNT 20 a 24 c 36 g 16 t 21 others
ORIGIN

ARI49027 Length: 117 February 21, 2002 16:59 Type: N Check: 1669 ..
Found using 'seq2-3' (pappu403.key)

1 CGGYCCGNNCAAYYYGNCGGCHNGSGYAGAGGTGNTAAGTCCGCCAAGTAAAGCT |-----|
44 49 55 60

61 GATCGAYGGNACANCAAGTATTGTGNACTTCACCCGTGTHGTHGC

4 matches found in sequence:
ari49028 ; TOIG of: ari49028 check: 7190 from: 1 to: 385
(from "mycobacter-ge.seq")
TOIG of: ari49028 check: 7190 from: 1 to: 385

LOCUS ARI49028 385 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 44 from patent US 6228371.
ACCESSION ARI49028
VERSION ARI49028.1 GI:15113619
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 385)
AUTHORS Nano, F.E.

TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 44 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..385
BASE COUNT 85 a 103 c 117 g 80 t
ORIGIN

ARI49028 Length: 385 February 21, 2002 16:59 Type: N Check: 7190 ..
Found using 'seq2-3' (pappu403.key)

71 ATGACGCCAACGCAATGATGTAGTACCTCATTTAGGCAACCCAGCGCTGACA |-----|
121 126

131 CTTTATGCTTCGGCTCGTGTAGTTGTGGGAATGTGGAGCGGATTAACAATTTCGACG |-----|
187

191 -|
192 ACGAGGAACACGCTGTAGACATGATGACGAATTTGAATACGACTACTATAGGAATTC

251 GAGCTCGTACCGGGGATCCTTAGAGTCTTCGCGCGGGTGCGCACATCAGGGCCA |-----|
292 297

311 GTGCGATCGCAAGCGCGGGTACCGGGCCATAGTCTTCAGATCGGCGGTGGACCGC |-----|
336 341

371 AGAGACCGGACGGG

2 matches found in sequence:
ari49029 ; TOIG of: ari49029 check: 6592 from: 1 to: 285
(from "mycobacter-ge.seq")
TOIG of: ari49029 check: 6592 from: 1 to: 285

LOCUS ARI49029 285 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 45 from patent US 6228371.
ACCESSION ARI49029
VERSION ARI49029.1 GI:15113620
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 285)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 45 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..285
BASE COUNT 47 a 101 c 87 g 41 t 9 others
ORIGIN

ARI49029 Length: 285 February 21, 2002 16:59 Type: N Check: 6592 ..
Found using 'seq2-3' (pappu403.key)

169 TCAGTACCCCGGCAATACGCCCGTATGNCACATCGGTCAGGGGTGCAAGCGTTGCGT |-----|
219 224

229 TGGCGTGATCGGCGGCGTGTATCGCCGTGATGCGCGCTGCTGTATCGGCGCGGTTCT |-----|
248 253

```
61 CGGTACACACCGAACCAGATGCGCGCGGCCCGCTGTAGAAAGGGAATCGC
79 84 116
121 CAGTGGCTGACTCGGTTTCATCCGACGCCAGTTGATCCTTTTGGCATGCTCCGCTAGTGG
121 142 147 165 170
181 CAATCGTCTATTGGGCTGTACTACCTCGCAATTCGAGTCTGCTGATCGGGCA
183 188
312 ACCATTGGCAAGTTTACTGCGGTGACGCCACCGACGAGGGCGCAGATGACGATGAGC
362 367
372 ATGCCGAGCACTACAAATCCCGTCGATGCTCGGCGAAGCTGCAATCGGTGC
372 377
486 TCCTCGGAGACACCATCACCAGGCGACCGCTTCCAGATCGGGCGCGCTGGAC
536 541
546 AATTCCAATCGCGGTTGGCCGATTTGCCACGAGAGAAATCGGCT
615 GCGGTGGGTGGGTCGGGACCGGCTTGCAACGGTTGTCGATTCACCTCAAGGATGTC
665 670 669 674
675 GGTGACTTCAAAACCAACATTTGGCGACGACGATCATTCGAGAACTCGGGCGCATT
696 701 705 710 699 704
735 TTGGACAGCCAGCTCAACAGGCTGA
...
3 matches found in sequence:
ari49024 ; TOIG of: ari49024 check: 1026 from: 1 to: 209
(from "mycobacter-ge.seq")
TOIG of: ari49024 check: 1026 from: 1 to: 209
LOCUS ARI49024 209 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 40 from patent US 6228371.
ACCESSION ARI49024
VERSION ARI49024.1 GI:15113615
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 209)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 40 08-MAY-2001;
FEATURES
1. .209 Location/Qualifiers
source /organism="unknown"
BASE COUNT 46 a 72 c 54 g 32 t 5 others
ORIGIN
ARI49024 Length: 209 February 21, 2002 16:59 Type: N Check: 1026 ..
```

```
Found using 'seq2-3' (pappu403.key)
...
19 GANATACGACGAGNDCCCGCCACGTTGCAAGACGGCGCAATCTTCCCTGTGCTACCC
69 74
79 GACTTTCGGCCCAACTTCACATCTTGCACATGACGCGCAABAABGCCGAGGTGCGGCG
135
139 ATGATGAAAGCCACCTTCCCGCGCGCTTCATCATCGACCGACGAGCTGACCGCTC
140 189 194
199 GACACCACTA
...
4 matches found in sequence:
ari49025 ; TOIG of: ari49025 check: 9709 from: 1 to: 167
(from "mycobacter-ge.seq")
TOIG of: ari49025 check: 9709 from: 1 to: 167
LOCUS ARI49025 167 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 41 from patent US 6228371.
ACCESSION ARI49025
VERSION ARI49025.1 GI:15113616
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 167)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 41 08-MAY-2001;
FEATURES
1. .167 Location/Qualifiers
source /organism="unknown"
BASE COUNT 24 a 50 c 60 g 30 t 3 others
ORIGIN
ARI49025 Length: 167 February 21, 2002 16:59 Type: N Check: 9709 ..
Found using 'seq2-3' (pappu403.key)
1 AGATCGTCAGTGAGCAGAACCCCGCCAAACCGCGCGCGAGGTGTTGTCAGGGGCTGA
3 8
61 AGCGCGTCTCGCGGACGCTGATGCTGCGGATCGGATCGGGCTTNGGCTCGCGCTG
62 67 91 96
121 TACTTCACGCGCGCATGTCGGCGCGGAGATGCTGATCATCGGGGA
132 137
...
1 match found in sequence:
ari49026 ; TOIG of: ari49026 check: 5999 from: 1 to: 221
(from "mycobacter-ge.seq")
TOIG of: ari49026 check: 5999 from: 1 to: 221
LOCUS ARI49026 221 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 42 from patent US 6228371.
ACCESSION ARI49026
VERSION ARI49026.1 GI:15113617
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 221)
```

```
1      TGTGGTAGCCTGCGCTCCATGATGCTTTCGATACGCTGTGACGCTTATTCATCAAG
      8      13      35      40      44      49      60
121    TCGTCGGCACTCCGTATATTCGCCGATTCCTCCCGCGGGAGCCAGCTGCTCGAGC
      65      87      92
121    TGGCTTGTGGGTATCGATGCGGCGAGCCGAGCGGTTTTCGAGATAGATTCAACA
      143      148
181    CCGGCAACGAGAGAGCGCCTTG
      194      199

5 matches found in sequence:
ari149021 : TOIG of: ari149021 check: 1112 from: 1 to: 319
(from "mycobacter-ge.seq")
TOIG of: ari149021 check: 1112 from: 1 to: 319

LOCUS      ARI149021      319 bp      DNA
DEFINITION Sequence 37 from patent US 6228371.
ACCESSION  ARI149021
VERSION     ARI149021.1 GI:15113612
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 319)
AUTHORS     Nano,F.E.
TITLE       Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL     Patent: US 6228371-A 37 08-MAY-2001;
FEATURES
            Location/Qualifiers
            source          1..319
            BASE COUNT      60 a 79 c 92 g 79 t 9 others
            ORIGIN

ARI149021 Length: 319 February 21, 2002 16:59 Type: N Check: 1112 ..
Found using 'seq2-3' (pappu403.key)

...

87      GATTNCCATGATATGTCGGTANCGTTCGCCATCATGATGCTTTCGACCGCTTG
      137      142      146
147      ACGCTTATTCATCAAGTCGTCGGCAGCTCCGTATATTCGCCGATTCCTCCCGCGGG
      151      162      167      189      194
207      ACCGACTGCTCGAGCTGCTTCGTGGGTATTCGATCGCGGAGCGCCAGCGCGTTTC
      245      250
267      GGAGATAGTTCAACACCGGCAACGAGAGACMG

9 matches found in sequence:
ari149022 : TOIG of: ari149022 check: 4804 from: 1 to: 263
(from "mycobacter-ge.seq")
TOIG of: ari149022 check: 4804 from: 1 to: 263

LOCUS      ARI149022      263 bp      DNA
DEFINITION Sequence 38 from patent US 6228371.
ACCESSION  ARI149022
VERSION     ARI149022.1 GI:15113613
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 263)
AUTHORS     Nano,F.E.
TITLE       Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL     Patent: US 6228371-A 38 08-MAY-2001;
FEATURES
            Location/Qualifiers
            source          1..263
            BASE COUNT      55 a 98 c 79 g 31 t
            ORIGIN

ARI149022 Length: 263 February 21, 2002 16:59 Type: N Check: 4804 ..
Found using 'seq2-3' (pappu403.key)
```

```
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 263)
AUTHORS     Nano,F.E.
TITLE       Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL     Patent: US 6228371-A 38 08-MAY-2001;
FEATURES
            Location/Qualifiers
            source          1..263
            BASE COUNT      55 a 98 c 79 g 31 t
            ORIGIN

ARI149022 Length: 263 February 21, 2002 16:59 Type: N Check: 4804 ..
Found using 'seq2-3' (pappu403.key)

1      GGTACTGTGCTCGATGAGACTCCGCGTCTCGATTTCAGACGAGTCCGACACGAGTGCACCGG
      8      13      42      47      50      55
61      CTCCACCAAGCCGAGAAAGCCACACGCGCGGAGCCGACACACAGTGCAGCGTGCCA
      78      83      95      100
121      GAACGACACACCGACACAGTGCAGCGGAAA

...

170      ATGCCAGTCTGCTGTGACCAAGTGGGACGCTCGCGGCTCGGAGACGAGTCCGAGACGAGTCTGCGGTG
      220      225      228
230      CCGGACGAGAGCGCTCGGCGGCGTGTGATGCGG
      240      245      253

14 matches found in sequence:
ari149023 : TOIG of: ari149023 check: 3323 from: 1 to: 841
(from "mycobacter-ge.seq")
TOIG of: ari149023 check: 3323 from: 1 to: 841

LOCUS      ARI149023      841 bp      DNA
DEFINITION Sequence 39 from patent US 6228371.
ACCESSION  ARI149023
VERSION     ARI149023.1 GI:15113614
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 841)
AUTHORS     Nano,F.E.
TITLE       Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL     Patent: US 6228371-A 39 08-MAY-2001;
FEATURES
            Location/Qualifiers
            source          1..841
            BASE COUNT      176 a 255 c 251 g 156 t 3 others
            ORIGIN

ARI149023 Length: 841 February 21, 2002 16:59 Type: N Check: 3323 ..
Found using 'seq2-3' (pappu403.key)
```

```

REFERENCE      Unclassified.
AUTHORS        1 (bases 1 to 164)
TITLE          Nano,F.E.
JOURNAL        Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
FEATURES       peptides
               Patent: US 6228371-A 32 08-MAY-2001;
               source    Location/Qualifiers
                       1..164
                           /organism="unknown"
BASE COUNT     22 a         53 c         44 g         27 t         18 others
ORIGIN
...
AR149016 Length: 164 February 21, 2002 16:58 Type: N Check: 7570 ..
Found using 'seq2-3' (pappu403.key)

...
65   NANTGANTMNCGCCGGTGC GGCGCCGTGCAGCGCTGACACTGTGCCATCCTCCCTGAGACGCTGCTG
                                     |-----|
                                     115 120
125   GGTTCGCCCGGGGAGNTCGAMANTCGGGCATCGCACACATC
-----
5 matches found in sequence:
ar149017 : TOIG of: ar149017 check: 7863 from: 1 to: 237
(from "mycobacter_ge.seq")
TOIG of: ar149017 check: 7863 from: 1 to: 237

LOCUS      AR149017      237 bp      DNA              PAT      08-AUG-2001
DEFINITION Sequence 33 from patent US 6228371.
ACCESSION  AR149017
VERSION    AR149017.1 GI:15113608
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 237)
TITLE      Nano,F.E.
AUTHORS     Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
            peptides
JOURNAL     Patent: US 6228371-A 33 08-MAY-2001;
FEATURES    Location/Qualifiers
            source      1..237
                        /organism="unknown"
BASE COUNT   30 a         69 c         89 g         40 t         9 others
ORIGIN
AR149017 Length: 237 February 21, 2002 16:58 Type: N Check: 7863 ..
Found using 'seq2-3' (pappu403.key)

1   ACGAGCGCAACGCGATGCGACCCTCCACCGGCGCGCAACGAGAGCGACTGTGCTC
             |-----|           |-----|           |---
--|          35 40 47 52 58
61   GCCGCGACCGCCGANCCGACCGTCTGTCNGCGANGCGACGACCGCAACCGTTGACCCAC
    63                      |-----|
                             95 100
121  NTTGTCACGACGACGCTGATSAATCAGGT
...
166  CCGTCAGCGGTTGTCGGGGTGCGGTTCTGGGACACCGTCACTGTGTTGGGCGCTNCG
                                     |-----|
                                     216 221
226  GCGCGNTGTTGGC
-----
5 matches found in sequence:
ar149018 : TOIG of: ar149018 check: 7268 from: 1 to: 371
(from "mycobacter_ge.seq")

```

```

TOIG of: ar149018  check: 7268  from: 1  to: 371

LOCUS      AR149018      371 bp      DNA      PAT      08-AUG-2001
DEFINITION Sequence 34 from patent US 6228371.
ACCESSION  AR149018
VERSION    AR149018.1  GI:15113609
KEYWORDS
SOURCE     .
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 371)
AUTHORS   Nano,F.E.
TITLE     Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
          peptides
JOURNAL   Patent: US 6228371-A 34 08-MAY-2001;
FEATURES   Location/Qualifiers
            source          1..371
                                /organism="unknown"
BASE COUNT      76 a      139 c      99 g      54 t      3 others
ORIGIN
AR149018  Length: 371  February 21, 2002 16:58  Type: N  Check: 7268  ..
Found using 'seq2-3' (pappu403.key)

...

7 matches found in sequence:
ar149020 ; TOIG of: ar149020  check: 2610  from: 1  to: 202
(from "mycobacter_9e.seq")
TOIG of: ar149020  check: 2610  from: 1  to: 202

-----
LOCUS      AR149020      202 bp      DNA      PAT      08-AUG-2001
DEFINITION Sequence 36 from patent US 6228371.
ACCESSION  AR149020
VERSION    AR149020.1  GI:15113611
KEYWORDS
SOURCE     .
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 202)
AUTHORS   Nano,F.E.
TITLE     Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
          peptides
JOURNAL   Patent: US 6228371-A 36 08-MAY-2001;
FEATURES   Location/Qualifiers
            source          1..202
                                /organism="unknown"
BASE COUNT      34 a      57 c      63 g      48 t
ORIGIN
AR149020  Length: 202  February 21, 2002 16:59  Type: N  Check: 2610  ..
Found using 'seq2-3' (pappu403.key)

```



```

...
135  ACTTACCTGTCTGTGGGCTCTATCCCGGTGACCAGGTCCGCTCCGTGGCGTCCGG
      |-----|
      185 190
195  TGGGTGAGATGCACATGATCGAGCCGCGTCTCGACGTCAAGATCACTATGTGGTGT
      |-----|
      223 228 235
255  CCAAGGAGCTCAAGGTGCCCGGTGAGTGCAGGCC
      |-----|
      260 265

3 matches found in sequence:
ar149012 ; TOIG of: ar149012 check: 9239 from: 1 to: 198
(from "mycobacter_9e.seq")
TOIG of: ar149012 check: 9239 from: 1 to: 198

LOCUS      ARI49012      198 bp      DNA
DEFINITION Sequence 28 from patent US 6228371.
ACCESSION  ARI49012
VERSION    ARI49012.1 GI:15113603
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.

REFERENCE  1 (bases 1 to 198)
AUTHORS   Nano,F.E.
TITLE     Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
          peptides
FEATURES   Location/Qualifiers
            source          1..198
                        /organism="unknown"
BASE COUNT  48 a 55 c 52 g 39 t 4 others
ORIGIN

ARI49012 Length: 198 February 21, 2002 16:58 Type: N Check: 9239 ..
Found using 'seq2-3' (pappu403.key)

...

20  AGCCAATACCTATGACAGCGTGTGGCAGCGCTGACACCGATCCGCTGCCAGTCTCC
      |-----|
      70 75
80  CCATTGTGCAGGTGACTGACGACGACGACCGGACAGAGTATCGATAGCGCGAAT
      |-----|
      130 135
140  GCCGGCTTGAGACCGGGTGAATTATCAGAACTTTCAGTCACGAGAGCGGTGATTTT
      |-----|
      182 187

1 match found in sequence:
ar149014 ; TOIG of: ar149014 check: 5059 from: 1 to: 210
(from "mycobacter_9e.seq")
TOIG of: ar149014 check: 5059 from: 1 to: 210

LOCUS      ARI49014      210 bp      DNA
DEFINITION Sequence 30 from patent US 6228371.
ACCESSION  ARI49014
VERSION    ARI49014.1 GI:15113605
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.

REFERENCE  1 (bases 1 to 210)
AUTHORS   Nano,F.E.

```

```

TITLE      Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
          peptides
JOURNAL    Patent: US 6228371-A 30 08-MAY-2001;
FEATURES   Location/Qualifiers
            source          1..210
                        /organism="unknown"
BASE COUNT  42 a 78 c 59 g 31 t
ORIGIN

ARI49014 Length: 210 February 21, 2002 16:58 Type: N Check: 5059 ..
Found using 'seq2-3' (pappu403.key)

...

53  AGCAGCGCGCGGTATGCTGTGGCGCGGCTAGACCAAAACTTTAGCGCAGCGCCGA
      |-----|
      103 108
113  GCCACCGACTTCGAAGCCCTCGCGCGGTGGTTGCGACATGGGT

1 match found in sequence:
ar149015 ; TOIG of: ar149015 check: 7536 from: 1 to: 255
(from "mycobacter_9e.seq")
TOIG of: ar149015 check: 7536 from: 1 to: 255

LOCUS      ARI49015      255 bp      DNA
DEFINITION Sequence 31 from patent US 6228371.
ACCESSION  ARI49015
VERSION    ARI49015.1 GI:15113606
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.

REFERENCE  1 (bases 1 to 255)
AUTHORS   Nano,F.E.
TITLE     Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
          peptides
FEATURES   Location/Qualifiers
            source          1..255
                        /organism="unknown"
BASE COUNT  47 a 93 c 69 g 42 t 4 others
ORIGIN

ARI49015 Length: 255 February 21, 2002 16:58 Type: N Check: 7536 ..
Found using 'seq2-3' (pappu403.key)

...

38  ACCTTCGACACGTTGCGCGGTCAACCTGCANTGAGCGCAACCTCTCGGTGCGTCAAC
      |-----|
      88 93
98  GAGTGTGCGCGGCTTCAACCGCTGAGACCGCGAACCCTGCAGCT

1 match found in sequence:
ar149016 ; TOIG of: ar149016 check: 7570 from: 1 to: 164
(from "mycobacter_9e.seq")
TOIG of: ar149016 check: 7570 from: 1 to: 164

LOCUS      ARI49016      164 bp      DNA
DEFINITION Sequence 32 from patent US 6228371.
ACCESSION  ARI49016
VERSION    ARI49016.1 GI:15113607
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.

```

...

2 matches found in sequence:
ar149008 ; TOIG of: ar149008 check: 459 from: 1 to: 207
(from "mycobacter-ge.seq")
TOIG of: ar149008 check: 459 from: 1 to: 207

LOCUS ARI49008 207 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 24 from patent US 6228371.
ACCESSION ARI49008
VERSION ARI49008.1 GI:15113599
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 207)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 24 08-MAY-2001;
FEATURES
source 1..207
Location/Qualifiers
/organism="unknown"
BASE COUNT 28 a 82 c 59 g 29 t 9 others
ORIGIN
ARI49008 Length: 207 February 21, 2002 16:58 Type: N Check: 459 ..
Found using 'seq2-3' (pappu403.key)

41 GCCTACGGGATCCCTGCTGCGACCGCCGCTATCCTCGCCCAAGTCTCGCGCGGCA 1-----1
91 96

101 ACAGTCCCGYCCGTGACAGACCTCAGGGCTNCTHCCGGGGGAGACAGACAGGTCTC

161 ACCGGYCCWNGYACCCCGACCGATCGCTGCGCGATTCGCGTGGCA |-----|
189 194

1 match found in sequence:
ar149009 ; TOIG of: ar149009 check: 9014 from: 1 to: 204
(from "mycobacter-ge.seq")
TOIG of: ar149009 check: 9014 from: 1 to: 204

LOCUS ARI49009 204 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 25 from patent US 6228371.
ACCESSION ARI49009
VERSION ARI49009.1 GI:15113600
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 204)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 25 08-MAY-2001;
FEATURES
source 1..204
Location/Qualifiers
/organism="unknown"
BASE COUNT 29 a 63 c 53 g 45 t 14 others
ORIGIN
ARI49009 Length: 204 February 21, 2002 16:58 Type: N Check: 9014 ..
Found using 'seq2-3' (pappu403.key)

93 TCCTGTACTNACGACCTGATCGGCATTTCACCAACCCGTGNGCAGACATCGCACCG |-----|
143 148

153 TCCGCGCCGCGYCGGTCGGAGCTGTGATCTTGTRKCGAGTAGAG

1 match found in sequence:
ar149010 ; TOIG of: ar149010 check: 4369 from: 1 to: 207
(from "mycobacter-ge.seq")
TOIG of: ar149010 check: 4369 from: 1 to: 207

LOCUS ARI49010 207 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 26 from patent US 6228371.
ACCESSION ARI49010
VERSION ARI49010.1 GI:15113601
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 207)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 26 08-MAY-2001;
FEATURES
source 1..207
Location/Qualifiers
/organism="unknown"
BASE COUNT 23 a 77 c 57 g 41 t 9 others
ORIGIN
ARI49010 Length: 207 February 21, 2002 16:58 Type: N Check: 4369 ..
Found using 'seq2-3' (pappu403.key)

86 NCCTTACAGCGGCTTGCGCGCTTCGCGGTAGACCGTTCGATGCGGCTTCGCGCGCTGCG |-----|
136 141

146 GTGTGCGTCTGCTGCGGCTTCGCGGTGCGCNCCTGCGCCGACCCG

4 matches found in sequence:
ar149011 ; TOIG of: ar149011 check: 4258 from: 1 to: 289
(from "mycobacter-ge.seq")
TOIG of: ar149011 check: 4258 from: 1 to: 289

LOCUS ARI49011 289 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 27 from patent US 6228371.
ACCESSION ARI49011
VERSION ARI49011.1 GI:15113602
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 289)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 27 08-MAY-2001;
FEATURES
source 1..289
Location/Qualifiers
/organism="unknown"
BASE COUNT 40 a 85 c 97 g 57 t 10 others
ORIGIN
ARI49011 Length: 289 February 21, 2002 16:58 Type: N Check: 4258 ..
Found using 'seq2-3' (pappu403.key)

VERSION ARI49004.1 GI:15113595
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 156)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 20 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..156
BASE COUNT 37 a 48 c 38 g 28 t 5 others
ORIGIN
ARI49004 Length: 156 February 21, 2002 16:58 Type: N Check: 8746
Found using 'seq2-3' (pappu403.key)

1 TCTTACAGGAGCGCTTCGCCAGACAGACGACTGTGACGACTGGNCGTCAAGC
12 17 42 47 57

61 TCACAAATGGCTTGTCCGATCTGTACRAGCAAGMTGAGTCCGTCGCCNMGCAACGCGA
62 99 104

121 CGAGATCATCGAGACCTACACCGTTGCCACGAA

1 match found in sequence:
ari49005 ; TOIG of: ari49005 check: 1942 from: 1 to: 123
(from "mycobacter-ge.seq")
TOIG of: ari49005 check: 1942 from: 1 to: 123

LOCUS ARI49005 123 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 21 from patent US 6228371.
ACCESSION ARI49005
VERSION ARI49005.1 GI:15113596
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 123)
AUTHORS Nano, F.E.

JOURNAL Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
FEATURES Patent: US 6228371-A 21 08-MAY-2001;
source Location/Qualifiers
1..123

BASE COUNT 11 a 39 c 40 g 24 t 9 others
ORIGIN

ARI49005 Length: 123 February 21, 2002 16:58 Type: N Check: 1942
Found using 'seq2-3' (pappu403.key)

42 CGGTGACCCAGTGGCGNTGTGGGGCTCGCTGACGGCGAGTGTGNCGAGCGTCCGCG
92 97

102 GTCGTGNCCTNTCMTCCCGCC

2 matches found in sequence:
ari49006 ; TOIG of: ari49006 check: 6851 from: 1 to: 823
(from "mycobacter-ge.seq")
TOIG of: ari49006 check: 6851 from: 1 to: 823

LOCUS ARI49006 823 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 22 from patent US 6228371.
ACCESSION ARI49006
VERSION ARI49006.1 GI:15113597
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 823)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 22 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..823
BASE COUNT 127 a 272 c 280 g 130 t 14 others
ORIGIN
ARI49006 Length: 823 February 21, 2002 16:58 Type: N Check: 6851
Found using 'seq2-3' (pappu403.key)

187 CGGGGAGKCGGGGCCGCCACACCGGTTCTGTGTATCGGACGGCGGCCGACCA
237 242

247 CACGTTGCGTTGCTGCACATCGCCGATCGGGGAGATTCTTCT

567 GCGCACAAAGCGCGTGCTAGCAACTTGGCGCGCTGGCGCGGACGAGGTACGCCACTG
617 622

627 CGAAATCCAGCCCGGCTTTTCGACGCGGGTTACGTCGTGGGGT

1 match found in sequence:
ari49007 ; TOIG of: ari49007 check: 1956 from: 1 to: 103
(from "mycobacter-ge.seq")
TOIG of: ari49007 check: 1956 from: 1 to: 103

LOCUS ARI49007 103 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 23 from patent US 6228371.
ACCESSION ARI49007
VERSION ARI49007.1 GI:15113598
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 103)
AUTHORS Nano, F.E.

JOURNAL Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
FEATURES Patent: US 6228371-A 23 08-MAY-2001;
source Location/Qualifiers
1..103

BASE COUNT 16 a 45 c 27 g 15 t
ORIGIN

ARI49007 Length: 103 February 21, 2002 16:58 Type: N Check: 1956
Found using 'seq2-3' (pappu403.key)

1 CTTCCGGCGGAGACAACAAGGTCTACCGGCGGACACCTGACACCTGATCGGCTGTG
30 35

61 CCGATCCCGGTGCGAGACCCGGGT

ORIGIN

ARI49000 Length: 170 February 21, 2002 16:58 Type: N Check: 2866 ..
Found using 'seq2-3' (pappu403.key)

...

11 AAGCGMACCCGGTCGGAGAGCCTGCACAGTTTACGCCGTGTCGCCGCGACGCTACTC |-----|
61 66

71 CTCGCTCGGGGGTCCCAAGCAGCCAGCAGCATCCTTGACCTTATTCGCAACGCG

131 CAATCCAGGCCAACGCCGAGGATCATCGACACCGACA |-----|
143 148

1 match found in sequence:
ari49001 ; TOIG of: ari49001 check: 6419 from: 1 to: 181
(from "mycobacter-ge.seq")
TOIG of: ari49001 check: 6419 from: 1 to: 181

LOCUS ARI49001 181 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 17 from patent US 6228371.
ACCESSION ARI49001
VERSION ARI49001.1 GI:15113592
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 181)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL peptides
FEATURES Patent: US 6228371-A 17 08-MAY-2001;
source Location/Qualifiers
1..181
/organism="unknown"
BASE COUNT 28 a 60 c 55 g 28 t 10 others
ORIGIN

ARI49001 Length: 181 February 21, 2002 16:58 Type: N Check: 6419 ..
Found using 'seq2-3' (pappu403.key)

...

10 CGCGGCTNACNCNGCGTGCCTTGACACCGCCACANTGACCTGCCCCGACGCCGAG |-----|
60 65

70 TGGTTCGCGCNCGGCGGCGCTTCGAAACCGCGGATTCGACACG

1 match found in sequence:
ari49002 ; TOIG of: ari49002 check: 9039 from: 1 to: 95
(from "mycobacter-ge.seq")
TOIG of: ari49002 check: 9039 from: 1 to: 95

LOCUS ARI49002 95 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 18 from patent US 6228371.
ACCESSION ARI49002
VERSION ARI49002.1 GI:15113593
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 95)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides

JOURNAL Patent: US 6228371-A 18 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..95
/organism="unknown"
BASE COUNT 16 a 34 c 29 g 11 t 5 others
ORIGIN

ARI49002 Length: 95 February 21, 2002 16:58 Type: N Check: 9039 ..
Found using 'seq2-3' (pappu403.key)

1 AGGTACGGTGGCAGACGCCCGTCGAATCGGWTGAGAAVGTCTCGKACACGCCA |-----|
14 19

61 GCTGGTCC

4 matches found in sequence:
ari49003 ; TOIG of: ari49003 check: 1732 from: 1 to: 283
(from "mycobacter-ge.seq")
TOIG of: ari49003 check: 1732 from: 1 to: 283

LOCUS ARI49003 283 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 19 from patent US 6228371.
ACCESSION ARI49003
VERSION ARI49003.1 GI:15113594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 283)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL peptides
FEATURES Patent: US 6228371-A 19 08-MAY-2001;
source Location/Qualifiers
1..283
/organism="unknown"
BASE COUNT 46 a 66 c 95 g 61 t 15 others
ORIGIN

ARI49003 Length: 283 February 21, 2002 16:58 Type: N Check: 1732 ..
Found using 'seq2-3' (pappu403.key)

...

72 AGCGGCTGTGCACCNACAGCAACGAGTTCGCTCGGCTCCGAGGTGCGATGACG |-----|
122 127

132 TGGCGCGCGAGGTTTGAGAGGTCATGCTGTTTCGCGGNTAGTGCCTCGATGATTGAT |-----|
133 138 158 163

192 GACGGAAGCNCNGTNGAANTTGGCNAATACGGCGTTGTGNGTCTGCNATNGACGRAT |-----|
224 229

252 TNCCTGCTGTCAGTGTAGNCGTGTGA

4 matches found in sequence:
ari49004 ; TOIG of: ari49004 check: 8746 from: 1 to: 156
(from "mycobacter-ge.seq")
TOIG of: ari49004 check: 8746 from: 1 to: 156

LOCUS ARI49004 156 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 20 from patent US 6228371.
ACCESSION ARI49004

FEATURES
source 1..395 Location/Qualifiers
BASE COUNT 65 a 111 c 149 g 70 t
ORIGIN
AR148997 Length: 395 February 21, 2002 16:58 Type: N Check: 5928
Found using 'seq2-3' (pappu403.key)

126 GGAGAGTAGCGGTGCGGGGTGAAGGTGCGAGATTGACTGTGGCGGTGCGGTGGG
176 181
186 GCTTGTTCGGCGCGGTGCGGGGTGTCCTCCGACAAACACCGGACGAGC
222 227
246 ATCCCGTCGACACCCGCAACCGTGAAGGCGGCGACGACCTTCTCCCGCATGTGGC
247
306 GCGCTCAGCCGATCAGACGCGGTGACGAGCTGACAAAGGTGACCGGCGGTCAACACCGCC
311 312 317
366 AA

2 matches found in sequence:
ar148998 ; TOIG of: ar148998 check: 9110 from: 1 to: 175
(from "mycobacter-ge.seq")
TOIG of: ar148998 check: 9110 from: 1 to: 175

LOCUS AR148998 175 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 14 from patent US 6228371.
ACCESSION AR148998
VERSION AR148998.1 GI:15113589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 175)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 14 08-MAY-2001;
FEATURES
source 1..175 Location/Qualifiers
BASE COUNT 34 a 49 c 47 g 29 t 16 others
ORIGIN

AR148998 Length: 175 February 21, 2002 16:58 Type: N Check: 9110
Found using 'seq2-3' (pappu403.key)

1 CCAGNCNCNACNAGTGTTCGCTCAATTCGCGCGGTGCGGTGCGTGTGCGAC
36 41
61 ATTCGACCGACCGCGTGAACCTGATCAGACGAACTGAGNAGATCAGCATNAAGCCG
121 GAGTGGCAGCAGCAACGCTGCTACTGGAATTTGGGAAAGCTGATGCTGAGCTGAC

137 142
6 matches found in sequence:
ar148999 ; TOIG of: ar148999 check: 7413 from: 1 to: 265
(from "mycobacter-ge.seq")
TOIG of: ar148999 check: 7413 from: 1 to: 265

LOCUS AR148999 265 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 15 from patent US 6228371.
ACCESSION AR148999
VERSION AR148999.1 GI:15113590
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 265)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 15 08-MAY-2001;
FEATURES
source 1..265 Location/Qualifiers
BASE COUNT 48 a 86 c 85 g 45 t 1 others
ORIGIN

AR148999 Length: 265 February 21, 2002 16:58 Type: N Check: 7413
Found using 'seq2-3' (pappu403.key)

1 GGGCTGATTTCGAGGCTGCTGATGCGGTGACGAGTACGAGGAGGATGCGCCAGCTGCTCAAT
43 48
61 ACCATCGGTGATTAACAAGGCTGACATGATGCTTGAATCTACAAGGCTGCGCTC
121 CCACCGACCCCGCGCCCTCGAGCTGGGGCTGTGCGGATCTGATCGCGCGACACT
132 137
181 TTTCGCGACTGTGCTGGGGGTGCGGGAAGAACCGACACGCGGAGCTCCCGAGTTC
191 196
241 CGGGTCGCCGTGCGCGGAGCCAC
244 249
250 255

2 matches found in sequence:
ar149000 ; TOIG of: ar149000 check: 2866 from: 1 to: 170
(from "mycobacter-ge.seq")
TOIG of: ar149000 check: 2866 from: 1 to: 170

LOCUS AR149000 170 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 16 from patent US 6228371.
ACCESSION AR149000
VERSION AR149000.1 GI:15113591
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 170)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 16 08-MAY-2001;
FEATURES
source 1..170 Location/Qualifiers
BASE COUNT 33 a 68 c 43 g 23 t 3 others

```

1 CCTGTTCACAGGTCNCTTCGGAACGAGCTTCTGATTCGATTCGNNCTCGGNCGTTCCCT
29 34
61 CGCACCGGTCCATGTTGATCAAGCTACGCTTCGCGGTGCTATGCTGCTGGTGCCG
87 92
121 CCGGTCTGTGTGTGTATTCGG
...

3 matches found in sequence:
ar148994 ; TOIG of: ar148994 check: 3708 from: 1 to: 120
(from "mycobacter_ge.seq")
TOIG of: ar148994 check: 3708 from: 1 to: 120

LOCUS ARI48994 120 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 10 from patent US 6228371.
ACCESSION ARI48994
VERSION ARI48994.1 GI:15113585
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 10 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..120
/organism="unknown"
BASE COUNT 24 a 42 c 37 g 17 t
ORIGIN
ARI48994 Length: 120 February 21, 2002 16:58 Type: N Check: 3708 ..
Found using 'seq2-3' (pappu403.key)

1 CAACGAGATCCACCCTGATTAGGAGGTGACGGTGGCAGCCGCGCTCGAATCCG
39 44
61 ATCGAAGTAACGCTCCGATGACGCCAGCTCGCGCGCGATGCGACCTGCCACCCGTC
69 74 80 85
121
-----
3 matches found in sequence:
ar148995 ; TOIG of: ar148995 check: 5910 from: 1 to: 162
(from "mycobacter_ge.seq")
TOIG of: ar148995 check: 5910 from: 1 to: 162

LOCUS ARI48995 162 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 11 from patent US 6228371.
ACCESSION ARI48995
VERSION ARI48995.1 GI:15113586
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 11 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..162
/organism="unknown"
BASE COUNT 26 a 65 c 48 g 19 t 4 others
ORIGIN

```

```

1 ARI48995 Length: 162 February 21, 2002 16:58 Type: N Check: 5910 ..
Found using 'seq2-3' (pappu403.key)

1 CGGCTTCACAGGTCGCGCAAGCAGCGCGCTCGTCCGAGATCTCCCAATGACGACG
43 48
61 CCGGCGCCCAAGACACCCCGGNTACCGTGCCTTCGTCCGCAACTCTGCAGCAACCC
63 68
121 GAGATCGCCACACNNCTACNCGCGGAGATGACGCTGTGCGG
124 129
-----
3 matches found in sequence:
ar148996 ; TOIG of: ar148996 check: 2585 from: 1 to: 133
(from "mycobacter_ge.seq")
TOIG of: ar148996 check: 2585 from: 1 to: 133

LOCUS ARI48996 133 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 12 from patent US 6228371.
ACCESSION ARI48996
VERSION ARI48996.1 GI:15113587
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 12 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..133
/organism="unknown"
BASE COUNT 14 a 34 c 45 g 28 t 12 others
ORIGIN
ARI48996 Length: 133 February 21, 2002 16:58 Type: N Check: 2585 ..
Found using 'seq2-3' (pappu403.key)

29 THGNCGHGCHGCAATCAGCTGCGCTTGCTTCGGTGCCGGGTGTCATCGCCCTTCG
79 84
89 ACCAGTTGTGGCGCTGCAACAGCATGAGCGGCTAGTCTATTCGG
98 103 117 122
-----
7 matches found in sequence:
ar148997 ; TOIG of: ar148997 check: 5928 from: 1 to: 395
(from "mycobacter_ge.seq")
TOIG of: ar148997 check: 5928 from: 1 to: 395

LOCUS ARI48997 395 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 13 from patent US 6228371.
ACCESSION ARI48997
VERSION ARI48997.1 GI:15113588
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 13 08-MAY-2001;

```

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 172)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 6 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..172
BASE COUNT 35 a 56 c 52 g 28 t 1 others
ORIGIN

ARI48990 Length: 172 February 21, 2002 16:58 Type: N Check: 8455
Found using 'seq2-3' (pappu403.key)

1 TCNCTTATCGCTTCAGTGGCATCTGCCCAAGACCGAATCTGGACCTATGACGGCCAGC
7 12
61 TGAAGATGGCCCGGACGAAAGGGCGTGGCAGCTTCGCTGGACCAAGCGGTTGCATC
82 87
121 CCAAGCTAGGGGAGCAT

5 matches found in sequence:
ari48991 : TOIG of: ari48991 check: 7396 from: 1 to: 232
(from "mycobacter-ge.seq")
TOIG of: ari48991 check: 7396 from: 1 to: 232

LOCUS ARI48991 232 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 7 from patent US 6228371.
ACCESSION ARI48991
VERSION ARI48991.1 GI:15113582
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 232)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 7 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..232
BASE COUNT 34 a 80 c 79 g 36 t 3 others
ORIGIN

ARI48991 Length: 232 February 21, 2002 16:58 Type: N Check: 7396
Found using 'seq2-3' (pappu403.key)

1 CTCTCGCGCCAGCNCGTCCCGCTGTCGGGATGNCGTACCGGTCTCTCAGCGCCAAGAC
44 49 50 55
61 GTTGACAGCTTAACGACGGCGGTTGGTGGCAGCGGTGCACTTGGCGGCCCAATGTGCG
71 76
121 GGGGCTGCTGAGTGGCGCCGCGTGGCGCTGTTGCAAAAGGACACGCTGGTGGCCGCCG
158 163
181 GACGGCCCCGATCGTGAAGGATGCAAGATCAGGTGAGTGCACCGCAATCGGAT

191 196

5 matches found in sequence:
ari48992 : TOIG of: ari48992 check: 3318 from: 1 to: 173
(from "mycobacter-ge.seq")
TOIG of: ari48992 check: 3318 from: 1 to: 173

LOCUS ARI48992 173 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 8 from patent US 6228371.
ACCESSION ARI48992
VERSION ARI48992.1 GI:15113583
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 173)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 8 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..173
BASE COUNT 38 a 51 c 44 g 37 t 3 others
ORIGIN

ARI48992 Length: 173 February 21, 2002 16:58 Type: N Check: 3318
Found using 'seq2-3' (pappu403.key)

1 GTTCGNCGGCTCAAAAGGTTGACGATGTCACGTGCGACGTGTCGCCGAGACCAAGT
22 27
61 GGATTGCGTGAAGACCTCAAAAGNCTCTACTGCAATGCTCAAGCCCTCAAGCAGCACC
96 101 111 116 114 119
121 GAAGGATTTGTCATCCTCGCTGCAGCTGTTGTCGAGCTTCCCATTTCCAC
155 160

2 matches found in sequence:
ari48993 : TOIG of: ari48993 check: 9534 from: 1 to: 223
(from "mycobacter-ge.seq")
TOIG of: ari48993 check: 9534 from: 1 to: 223

LOCUS ARI48993 223 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 9 from patent US 6228371.
ACCESSION ARI48993
VERSION ARI48993.1 GI:15113584
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 223)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides
JOURNAL Patent: US 6228371-A 9 08-MAY-2001;
FEATURES Location/Qualifiers
source 1..223
BASE COUNT 30 a 66 c 67 g 49 t 11 others
ORIGIN

ARI48993 Length: 223 February 21, 2002 16:58 Type: N Check: 9534
Found using 'seq2-3' (pappu403.key)

301 AATCCAAAGGGCTGACCAACATCCGATTGCGTGTGCGATGACAGCGGAAAA
361 CCCGACTTCCGGATTGCTGCTGCTGTCGCCGACGACAGTCCGGGGGTGTGCTACGAA
384 389 390 395
421 TTCCGGCTCCGAGCTCTGCTACACCCGCTTCGGCGGCGCATGATCCGATCCGGG
432 437 450 455 467 472
481 GTGTTCAACACGAGCGCGCTGGGTACGCG

13 matches found in sequence:
ari148988 ; TOIG of: ari148988 check: 8502 from: 1 to: 512
(from "mycobacter-ge.seq")
TOIG of: ari148988 check: 8502 from: 1 to: 512
LOCUS ARI148988 512 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 4 from patent US 6228371.
ACCESSION ARI148988
VERSION ARI148988.1 GI:15113579
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 512)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 4 08-MAY-2001;
FEATURES
source location/Qualifiers
1..512
/organism="unknown"
BASE COUNT 85 a 155 c 165 g 100 t 7 others
ORIGIN
ARI148988 Length: 512 February 21, 2002 16:58 Type: N Check: 8502 ..
Found using 'seq2-3' (pappu403.key)
...
105 TTGCAGTGTGNGCAGCCGAGATTGGAAGTTNCCGACATTACCGTTGCCAGCTTCCC
155 160
165 TCGCCGACGTTTCGCCCAAGCCAGGTGCGGACACGCCGGTGATGTGCTGGGGGCAATGA
170 175
225 NCGGGCTGCTGGCCCGCGCAATTCCAAAGCGTGATCGGACGCGTGTCCAGACCGGG
253 258
285 CCGCTCGCTAGCGTGACCGAGTCTTCTGAAATTGCGGATCAGCAGCTGACCTACCGCG
320 325 344
345 ACGCTAACGCCACCGCCCAACCGGTAACGCCCGGGTGTGGCCCGCGCGGCTCGGCCCG
349 350 355 392 397 404
405 GCGACGCTGTGGCATCATGTGGCTAATCACCACAGACAGTCTTGCGCATGCTGGCCA
409 412 451 456

410 415
465 CGGTCAAGTGGCGGCTATGCGCGCATGCTCAACTACACACGCGCG
476 481 482 487
7 matches found in sequence:
ari148989 ; TOIG of: ari148989 check: 846 from: 1 to: 456
(from "mycobacter-ge.seq")
TOIG of: ari148989 check: 846 from: 1 to: 456
LOCUS ARI148989 456 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 5 from patent US 6228371.
ACCESSION ARI148989
VERSION ARI148989.1 GI:15113580
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
AUTHORS Nano,F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
peptides
JOURNAL Patent: US 6228371-A 5 08-MAY-2001;
FEATURES
source location/Qualifiers
1..456
/organism="unknown"
BASE COUNT 63 a 121 c 165 g 98 t 9 others
ORIGIN
ARI148989 Length: 456 February 21, 2002 16:58 Type: N Check: 846 ..
Found using 'seq2-3' (pappu403.key)
...
130 GCGCTGATGGCGGAGCGAKTGTGACCGAAGGAAATGCTCAACATCCAGGGGTGGG
180 185
190 GCGTGTGTATCAAGCGGTGACCGTCAAGAGCGCGGTGATGATGATGACGCGT
220 225
250 GCTGCTCATCTACTGATACCAATCTGTTGCTGATCTGCTGATGCGGCCCTGGA
308
310 CGCGGNNGATNCGCTATGCTGAGACACAGGGGGTTCGCTGATGCTGCGGGGTG
313
370 CGCGGCGCTCTAANTCGTATCGCGCGGCTGAGACCTGTGATCTTGTGCTGGCGG
374 379 398 403 428
430 CGTTCCGCTGCTGTTTACCGCAGCGG
433 437 442
2 matches found in sequence:
ari148990 ; TOIG of: ari148990 check: 8455 from: 1 to: 172
(from "mycobacter-ge.seq")
TOIG of: ari148990 check: 8455 from: 1 to: 172
LOCUS ARI148990 172 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 6 from patent US 6228371.
ACCESSION ARI148990
VERSION ARI148990.1 GI:15113581
KEYWORDS


```

DEFINITION Sequence 1 from patent US 6228371.
ACCESSION ARI48985
VERSION ARI48985.1 GI:15113576
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 267)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 1 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..267
/organism="unknown"
BASE COUNT 56 a 77 c 74 g 60 t
ORIGIN
ARI48985 Length: 267 February 21, 2002 16:58 Type: N Check: 7728
Found using 'seq2-3' (pappu403.key)
...
49 GCGAGTGGCCCGACGAATGCTCGGTCTTGATTTCTTGATCTGGAAGACGTTGACC
|-----|
99 104
109 GAGTTGTTGGCACCAGGTTAGGCATCAGTCACGCTTGAAACAAGA
|-----|
259 264
209 TACGTTCCCGGACGACGACACACACCGAATTTGCGTGTGTCGCCCAAGCTTCG
|-----|
259 264
-----
10 matches found in sequence:
ari48986 ; TOIG of: ari48986 check: 502 from: 1 to: 487
(from "mycobacter-ge.seq")
TOIG of: ari48986 check: 502 from: 1 to: 487
LOCUS ARI48986 487 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 2 from patent US 6228371.
ACCESSION ARI48986
VERSION ARI48986.1 GI:15113577
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 487)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 2 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..487
/organism="unknown"
BASE COUNT 89 a 147 c 152 g 96 t 3 others
ORIGIN
ARI48986 Length: 487 February 21, 2002 16:58 Type: N Check: 502
Found using 'seq2-3' (pappu403.key)
...
1 CTGATACGACGCGGCAAGACTACGACGAGTGGACAGAAATTCATGCGCGCTCATC
|-----|
8 13
13 56
61 GGAACGACGTCGCCGCGCTGTTTCTGACGACACGATGATGCTTCATTTCGCCGCTC
|-----|
76 81 84
91 96
-----
121 ACGGCTGTTCTGACTCCCTTGACGACCTGTTGGCCAAAGTTGGGTGACACAACGAGAT
|-----|
142 147
181 TACGTGATTCGCTGCT
...
199 GCCGACTATGAGTTCACAGCCGCCATTACGCTGTGCGGTATGCTGCGTCCGACGCGCTG
|-----|
249 254
259 TTCTACTACAAACAGGCGCGTGGCAACAGCGCGCTTACCCGAC
...
314 AATCCTGTGTCAGAGTTGACGAGAGTGGGTCCGGAGTTACAGCGCGTGTGCGCGCGGTC
|-----|
364 369
374 GATCGGCGCAGCGCTGCGNTAACGCCGACATCATCTGTCGACGCTTTCAGGAGCAACT
|-----|
394 399
414 419
434 GGCATTTGGCGGCTGCTTACTCCGACAAAGTGACAT
...
-----
8 matches found in sequence:
ari48987 ; TOIG of: ari48987 check: 4808 from: 1 to: 511
(from "mycobacter-ge.seq")
TOIG of: ari48987 check: 4808 from: 1 to: 511
LOCUS ARI48987 511 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 3 from patent US 6228371.
ACCESSION ARI48987
VERSION ARI48987.1 GI:15113578
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 511)
AUTHORS Nano, F.E.
TITLE Mycobacterium tuberculosis DNA sequences encoding immunostimulatory
JOURNAL Patent: US 6228371-A 3 08-MAY-2001;
FEATURES
source Location/Qualifiers
1..511
/organism="unknown"
BASE COUNT 93 a 155 c 165 g 95 t 3 others
ORIGIN
ARI48987 Length: 511 February 21, 2002 16:58 Type: N Check: 4808
Found using 'seq2-3' (pappu403.key)
...
58 GACGACACGCTTCGCGGTCTTGACAGTGTGATGCCAATCTGCCGCGGAGAGCTGCCG
|-----|
108 113
118 GATGGCGCGCGGTTCACCGAGAGCTGTGACACAGACTGGCGTGTCTCCGGGACACTACT
|-----|
160 165
178 CCCAGGTGCGTCAAGGACCGTCAAGGTTCAGTA
...
241 GATCCCAATGTATAGGCGGTGACACGCAATTCGCCCAAGATGGTCCGACGACGAGTTCAC
|-----|
291 296
```

```
...
324098 CGCATTCGAGCCTCCGAACAATTTTATTAGAAATTCATCATGCGCCAAATCCGCCGTA 324148
      |-----|
      |-----|
      |-----|
324158 CTTAATTAACGGCTGATTTATCTGAGAGGATTTAGTGTCTTCTTCAGTTTGAATCAGTCA 324183
      |-----|
      |-----|
      |-----|
324218 AAGACGACAGTTTTCGAAAGTCTCTTTTCAACGGGATAAAAAAGTCTGGAATAATACG 324274
      |-----|
      |-----|
      |-----|
324278 TCTGATTTGACCGTTTATTGTGATCCAGCTCCCATCGTGCACGTGGCCCTTTCATTTAATC 324335
      |-----|
      |-----|
      |-----|
324338 GTTAAGTTACCCGTTATAGTCACTTCCCATTTATCACTAAAGGCTCGGTAAAA 324340
      |-----|
      |-----|
      |-----|
324457 AAGATTCCCTCGCGGGAACAATAATCTGCCGCCAATCCAGCGTAATACGTTTCAT 324507
      |-----|
      |-----|
      |-----|
324517 GGCCTTATCCGCTGAGGTTATCGTTGCCAGAAACCGCTTGTCCGGTTTTTACGG 324536
      |-----|
      |-----|
      |-----|
324577 CTATCTTCTGAAAGAATTAATAAGATCGTTTAAACGTAATTTTCCGTGATGAACAACG 324641
      |-----|
      |-----|
      |-----|
324697 GATGAGTAAGGAGGGCTGATTCGCTGTTATACGATAGTTACAGAATGTGTAATCA 324730
      |-----|
      |-----|
      |-----|
324757 GGTAACTGGCAAGCTTTTGGCTAAAGTAGCATTTGCCCGCGAGAGCGCTTGCATTCAT 324800
      |-----|
      |-----|
      |-----|
324817 TGACACGCGCTTCTTGATCCCTCATGGCATTAATAAACGCGCTTGATATGAGGCCAGTTTT 324851
      |-----|
      |-----|
      |-----|
324877 CAATGTCCGGGTACGGGTAATCTTCAGGCG 324880
      |-----|
      |-----|
      |-----|
...
324975 TTTTATCCGCTTCATCATATCCAGAAGATATATAGACCTGGAAGATATCTCTCAGA 325025
      |-----|
      |-----|
      |-----|
325035 GAACGTTGAATATATATATTTCTGTAAATTCGACGATTTTAAAGTGTCTGCAAG 325071
      |-----|
      |-----|
      |-----|
325095 AGAGCTGTCCCTGCAAGAAGTTTAGTATCAACAATAATGTATGACCATGTAATACTA 325101
      |-----|
      |-----|
      |-----|
325155 TG 325155
      |-----|
      |-----|
      |-----|
...
```

```
325493 ACTAATAATGCATATAAAAAATATTTCGTGTAGTCCCTTTCGTATGTAATAACCTCTTG 325543
      |-----|
      |-----|
      |-----|
325553 TTTTTCCTCCACACACCCTGTGACAAATTTTTTACTGCAAAAAGACGAGTTTGTACAGGC 325640
      |-----|
      |-----|
      |-----|
325613 TTGTGCCAAGACATATTCGACGAATCAGCGACGCGGCAAGAAGATGACTGTGTCTTT 325640
      |-----|
      |-----|
      |-----|
325673 TTTTGCTAACGGAAACGGAGAT 325640
      |-----|
      |-----|
      |-----|
...
325790 CGGCTGAGTTTATAAAATATTTCGAGACATATCTTCCATCGTAACGACGCTTAACA 325840
      |-----|
      |-----|
      |-----|
325850 AAATATAGTTGCGTTAACACCAAGTTGAATGATTTAATTTCTT 325840
      |-----|
      |-----|
      |-----|
...
326323 AACGGCAATTGTAGTGAAGACAGTCGCAAAATGGCTATTCCGCTGACACACACGTTAAT 326373
      |-----|
      |-----|
      |-----|
326383 TCCATTCGACTTTTAATCAATCCGATGGGGTTTACATGAACCTTTTAAATATACAG 326469
      |-----|
      |-----|
      |-----|
326443 CAATTGACGACATCGTATTCTCCGGTACGCTGTGGAGGTGTTCTCCTCAGTAGCGG 326469
      |-----|
      |-----|
      |-----|
326503 GCGGTGCGGTAACACACGGTGG 326469
      |-----|
      |-----|
      |-----|
...
326671 AGGCTCAGATGACAGCTCAATCGATCTGACCCAAAGCTGGCTTGTGAACAGCGCCACTC 326721
      |-----|
      |-----|
      |-----|
326731 TTGATCAGTGAACGGTAAGACTCTCATATGACAGTTAAACAAT 326721
      |-----|
      |-----|
      |-----|
...
327131 CTGCTAATCAACCCATTGCTTTGACGAAGTTGATTTAAATTTTCCCGGACGATCGGG 327181
      |-----|
      |-----|
      |-----|
327191 TGAATAATAGTTGTACTGTCTTGATGAGACATCACTTCATTTATC 327181
      |-----|
      |-----|
      |-----|
...
327467 CGCAAGGGGCGGTGATCCATTTTATGCCAAATTTGGAACGGCATGTGAGCTCAGCA 327517
      |-----|
      |-----|
      |-----|
327527 CCCAGTCATCACCCGTAGAATGAATGCCCCACAGAATGTTGTTCGGGGTAACACT 327517
      |-----|
      |-----|
      |-----|
327587 ACTCCAGCAAAAGCGTTAATAGCGGCAATGTCAAAAACGGCAAAATAGCTCAGTAAGA 327597
      |-----|
      |-----|
      |-----|
327647 TGCAGT 327597
      |-----|
      |-----|
      |-----|
...
2 matches found in sequence:
ari48985 : TOIG of: ari48985 check: 7728 from: 1 to: 267
(from "mycobacter-ge-seq")
TOIG of: ari48985 check: 7728 from: 1 to: 267
LOCUS ARI48985 267 bp DNA PAT 08-AUG-2001
```

319884 |-----|
AACGTTGCATACCAATAGGTGATAGAGGCACTGCGCAAGCTATTCTCAGGCAGAA
319884

320145 GGGCGAGCAAAATGCAAGATATGCTGTCAGTCAGGTGCTGATGTTGAGCGTTTGA
|-----| |-----| |-----|
320195 320203

320205 ---|
CGATTATGCGCATCCAGCAAAATAGTTCGATTGGCAACCGCTGCTGAATATCA
320208

320283 TTTGGGCGGAGCGCTACTGCGGAGTAAGTCTCAGAGCTTGCAAAACCTGCGCTTTCC
|-----|
320333

320343 GCTGCGTTGCTGAGATCGAACCCTAAATCGTGCCCTGCGCTGCAA
|-----|

320524 CTCTCAACCTGGCGCGTGGTTCATGTCGTGAAGATGACCTGCTGCGGCGCTGGAT
|-----| |-----|
320574

320584 AGCGGCAAAAGTTAAAGCGCAATGTTGATGTTTTTAATCGTAACCCCTTACCGCTGAA
|-----| |-----|

320644 AGTCCGCTCTGCGCAACATCCACGCGTGAAGATACACACATATGTCGCGCGATTAACCGT
|-----| |-----|
320670 320686

320704 CCCGCTGAAGCTGTGAGATACATTTCTCGCACTATTGC
|-----|

321133 GATCTGATTATTGCGGTTTTTCATGAGCGGTTTTTGCACACATGACAAAGCAAGCAAT
|-----|
321183

321193 |-----| |-----| |-----|
ACACAAGCGATGATGCCACTATTCGCCAGCGCAATGTCATATATAAGCCATCCCGGA
321198 321205 321214

321253 AATCCGAATATGAGATAGATGTAAAGCCGTTGCTGAAGACGCGCAACATCAGGTG
|-----|
321312

321313 ---|
GGCGTGAATAATCAATATTTCTCATTTTACACTCAGTAAGGCGAGTAAGACA
321317

321772 AAAGCTGTGATATGACGCAAGTCTGCGCGATTACATGCGTTGTTATCGGCGATGAT
|-----| |-----|
321822

321832 GTGCGTGCGCGCATATCGTAGCGGATGGGTTAGAGACGCAACGGA
|-----|

322370 CTGGGCTGTAGCGCTACTTTGCTGCGCCGCAAGGTGGGCTGAAGGCGCTGCGATCTCC
|-----| |-----|
322420

322430 GCTGCAACTCTGCTTAGTGGCGGTGTGGGAGATGCTATTTATTACGTAAGTGCACAG
|-----| |-----|
322459 322489

322490 ---|
GGCCACATCTGGAATTTCTCAGTTATGTATCAATACCGGATATTGCGCGTTTCTGATGTGT
322494

322550 ATTCAGGCCAAACACCTGCTGCTTTTCATTGTTCCGGGAGCGTTATAGGCCCATGCCG
|-----| |-----| |-----|
322588 322609

322610 ---|
ACGTTTGGCGGCGAGGTGACTGGAACGTTGTTACCTTCGCTGCGGCTTGCGCTGATA
322614 |-----|
322654

322670 TTTGGTTACGCAATGAATAAACACGCGCTGTGGCTGCGCG
|-----|

322970 GCACAGCATACGGGTTGTTGAGGTGTAAACCCACTTCCCTTCACAGCAACGCTGTA
|-----|
323020

323030 GTCAATTAAGCGGAAACCCCGGCTGAACCTTCATAGGAAGTATGCTGTCTGCTGT
|-----| |-----|
323073

323090 CACGAGGAGAGATCTCACCGGTAOTCATTTGACGAGCGCAGATTTCACGCAATCTG
|-----| |-----|
323123

323150 ATCGAGCTGTATTGCTGTGCGCACCGA
|-----|

323315 TTGTGCGGACAGATTAATCTTGCGCATTAAGCAGGTTTGTAAAGCCCTGCGCTCAG
|-----| |-----|
323365

323375 CGGTATAAACGACGAGATCTTTCATAGTCCGTGACCAACGATGCGGCTTGGCGG
|-----| |-----| |-----|
323415 323423

323435 AACACAGTGGAAAGTTACTTGCCGGTAGGCTTAAATTGCCCCGGTTGCTCTGAAT
|-----|

323495 GTTGTATACCGAAACAAAGATCTTACCCGCTGCGCGCTGACAGATGGTCAAACTTTGTA
|-----| |-----|
323526

323555 GCTCTGAGCAGAGCATTAATGTCCG
|-----|

323600 CGGGCGGTTAAGCATCCGCTCAGTACATGACGCGCAACCAAGAAATAACGCGTGCAT
|-----| |-----|
323650

323660 GATTATTTTCTTATGAACTGGGCTTAAAAATCGTTGAGTTAT
|-----|

323816 ACCATGCGGCGGTTTACCGGTATTAATATTCGACAGTAGCCCACTAATAATTGTGAC
|-----| |-----| |-----|
323900 323929

323876 TGGATGCCCTGAGTAAGTTATCTAAGCTGAGGCTGTTCAATACCAAAAGTCATCTGTA
|-----| |-----|
323908 323936

323936 TAGCTCGGATCTTTTAAAGAGTTTGGCGCTGAGGCTATTTAATAAAAAACGCGCATTA
|-----| |-----| |-----|
323968 323986

323996 TTTGGGTTACACCAAGTTTGGATTACGGAAGTGAAGTCAATG

316882 CGGACGTGCTGAAGCGTATCTGTCGTGAGTGCAGATATTTCCGGCGTCTTGCCGT
316742 GTTTCGAGAGCCCGTAAAGACAGTTTGGCCGTATGTCGGCGTGAATGATGTCACCCCT
316802 GAACTGATGTTCGGTCACCCGGGAAACGAGTACGACAGATGACCTCAGACGCCGCTGTG
316862 GATGAGGCCATTCTCATCAATGCATACATCTTCACCGAAACCGGTACGCCGCCGCTGA
316922 CCTGTACGTGTTAAACCATTTCTTAACCACTTTTCAGAGAGATTTTATCGTGTGACACAA
316982 AC
317011 ATCAGACCGCCCTGTGTGGGGCTTCCCTGCACAGTAGGCCCTGTCTTGGCGCTGTC
317071 TGGTGCAAGAGGGTAAACCGCCCTGCANTACCTTGCCGACCGCCGGTATCAGAGCCGCT
317131 TCAGGACGCGAGATGCATACCATCTGGACAGGCCCTTCCGCTGTGATGAACAACAT
317638 TGTCCCGCCTCTCAGTTAATAAACAAGTATGATATCTCCGGGCTCGCCGGCGACCGGC
317698 CTGATGACCCCGGACATTACAGAACGTAATAACATTACCTGCG
317746 AAGTATCCGGAGCGAAATGATGAACATGGCCCTGCACACTGAGACCCGACACCGTTAACG
317806 TACAGGCACTGAAACATGGGGCGCATTTGCTTGACGTGATGTAATTGAGTCTCGCTGAAC
317866 TGATTACATGTGCTGCGATTAACGGGCTA
317953 CTACCCCTTACCGGCACTACGCTGCAGTACCGCACATATACAGGAAACGACACCGCTGGC
318013 TGTACTCGCTGTACACACAGACAGTGAATTCGTGTAATCAGAAATG
318077 CGGCTATCTGTTGCGTACCGATGCGTGTGTCGTAACCCGGTTCGCGCTTAAACGCGCTGGG
318137 TCTGTCAAAAACGTTCCGTCGTGCTGTGCTACACTATCCGGCGTTATGCGCTAGCTCT
318146 318154 318162 318178 318186

318197 CATTCATCTGATGTCGCGCAGCGCTGAATGCTTGCAGGGTTTACCCACTTTCGACTGTGAAC
318257 AGGAACACATGAAATTCATTAAACACGGAACGCGCTGGAATTTCTGATTCGTGGCTG
318317 CAGACAAATATGCACTGCGGATCGGGAATTTATCTTTGACAAAGCATGAGATTAACGAT
318377 TCAGCAGCACTGTGGCCCTGTATCGAACAGGCCG
318441 ACCTTCTGCACCGAAGACCGGTGAAGCTCAGTCACAGACTTTCCTCTGACGCCAGCC
318501 TGCACAGGGCTAGCGTTTTCATTATCTGCAAAAAGGAATATGATTTATGTCAGATATCA
318561 CAATCTCCCGTCCGGAGTGTCAACGGGCATACGAGCTTATCTGTCACACCAACGATTC
318621 GCCAATTTCTGGCTGTACGAAGAGTACACT
318796 GAAACCATTTACCCGGAATCTTGATCGGAATCTGGCTGACCTGATGCAACGTTCCGG
318856 GATGCTGTATTAATGATGATACGACAGCCCGTGTATCATGTGACCGG
318957 TTGAACAACTGCACGAGATAAGATGAGGTGTTTGAAGCAGAGGGGTGATCAACGCTTCA
319017 GAGGCTGACCTGGAATTACAAACCAATTGCCCTGTAAATTTGGCAGTAAATTAATGCG
319077 TCACAAATCTGATGATGGAGCCGGTGGGATTTTCATCTTATACACCGGACA
319583 ACACACTTTCCAGGCGTGTCTCCTTCAGCCACTCGGACACCTACCAAAATTTGCTCTCG
319643 TCTTGTGGAACGGGCGCTAATTTAGGAAATCATGACCTGAGGTCAACAACTTTTATA
319703 AAAAATCGC
319764 TTTTGGCCACGACGAGGATTTGTTATGCTGTGTGCTCTTTGTAGATCATATGATAGT
319824 GCGAATAAATTTTCGACAAACGCTTTTCGGGAGTCAATGATATCATCTTTATTCACCC
319841

```
313432 GAGCATCTGGGGAATATACAGTGTGCTGTGGTAAACAGATTGCTTT
...
313504 CAGGCAATGAACCTCTGCCAGATCCCGGAGCTGACAGACACTGCTTACGACCTCTTT
313554
313564 CGCCATATATCGACATCAACACGGGTTGACAACTGTGATGCTCAGCCGTTGTGTCAAGAT
313624 GAGCCCTGACACGAGCTGATGATCAATTACTTACTGATGCTGTGGACAGCGGCTA
313635
313684 CCGCTGATGTGTGCTGTACCCACAGAAAGGGGCGATGAGATATTATTGACTGCATTT
313695
313744 CGTCTGAACCTGGCTGTGCGAACCAGCAAAAGCCGTTTTCAGCATATCCAGAGCCTG
313774
313804 CCCCAGGACAGCAGAAAAGAGCTCT
...
313891 GGTGATGTGTGTATGGAGATGGCGAAATATGACAGGAAATACCCGTGTGACGCTCTG
313941
313951 GTCGATGCCGATGCATATGTGTATGTGACGATCTGAGTACAGGTTCTTTATGGGAT
314009
314011 ---|
314014 CGCAGGCTTTTCCGACACCACAAATCCTTGAGATCCGTCGCCGGAAAGGCTG
...
314075 GTGATGTGGCAACAGCGCGGTGCTGCTCAATTTCATCGGCACATACGAGCGCTGGC
314125 314132
314135 GTCAGCAGGGCTATGAAGACAGATGCTGGCGATGGAGCATATCCGAAACCGCTGGCAG
314137 314163
314195 CACGTACGGCAGTGAACCCGGTCAG
...
314250 GGAAGAGCAGATTGTGCACTGAGAAACGGATGGCCCGATTAAATAAGAGCGCCCA
314300
314310 AGAAGCCGACCGCCTGTGTGTGTTTCTGTCTGTGTATTGAAAAATATTTTCA
314313
...
314438 GTGAACATATTTCCGGGGGTATCCGCAATGCTGAAGTCACTTGCATGGCGATGCCG
314488
314498 GAATACAGCTTAATACGTGTGCGCCAGTGAAGGCTTCTCAGACA
...
314559 TTCACCCATCAGGCGGCTTCTCCGGGAGACGGGCTTTTCATTATATGCTTAAT
314559
314619 CATTATPAGGAGCAAAATCATGATTAAGCCAGTGTGTTTGTATGGAACGACATAGC
314651
314679 TCGTATCGTCCGCTGACACAGAAAGACTGATGCCCATGTACCGATATTTTGGGGA
314684
314739 A
...
315056 ACCGGGAAATGTGTGACAGAGTCATTTGAAGGGGCTTACGAGTGTGCGGCTTTTG
315106
315116 ACCGATAGAGAGACAGCTGATGCCATGACAGTGTGCTGTCCGCCACCGGACGCC
315147
315176 AGGCGTGGCAGACGGCGCACTGACTTACGGTTATGTGACGAACATCAGCCGTCACCA
315177
315236 CCGCGCAATTTCTGACACACGAGCGCGGAGATTAAGGTAAAGACCTGTGAGTCTT
315257
315296 ATCAGACCATCCAGAG
...
315382 AGCATTTATACCGCATTTAAGCTCAATCCGCAATTGTGGTGATGGAGAAACGCTCTG
315432
315442 GAGAGCATGGCGTGTATGCCGTTTCACTGAAGTCACTCCTGACTC
...
315767 ATCTGTACACCTGGAACAACGGCGGCAATTTATGGACCGGAACCGGATGACGATG
315817
315827 ACGAACATGGGTACTGTTCATGTCATGAACGGCAACCCGCTGA
...
315956 CGGTCACTATTACCGGCTGCGGATTAAGCCCTGACAGCATCCGGAAATGACGGCCATTA
316006
316016 TGGCATCATTTGACTGAATGAGGACAGGAATGACAGACTTTCCTTTCTGCCGAG
316076
316080 AGATGACGCCCGGAGACGCAGTCTCATTTCTGGGGGCCCTGCAAAACCTGAGACGCCATC
...
316371 TCACCCGTCCGTTGAAGTACACACCAAGTAAAGCAGACCGGCTTATCACCGAAGCTCGT
316421
316431 ACAAGCACTGGCGCTGTGTGATATCCGGGTGCCGACATCTGATA
...
316622 TCACCAACAGCATCTGATCCGCTCTTTTCTGTTCTGTACGGTAAATGCTGCGCA
316672
```

```
309916 CTGAATTATTGAATTATTAAGGTGAATGATGCACGAAAAAAGCAATCGCCCTCTTGTG
309960
309976 ATGAGCGCAGACGGCTTTTGCACACTGAATATCAATCTGCT
...
310041 GTGTATCTGCAGATAAGAAAGAAAGGGCTGCTTTTGTATACGGAAAGCGCGTGA
310091
310101 AGAATTGAGAGAGCTGGAGGGCGAACAATCAAAACCGCCCGCAGGG
...
310150 TGGTGGCTGGCTGTGTGCGCAGATGAAGCGGGCAATACACCACTCAAGCGCATTTG
310200
310210 TGGGGCAGAAATTTGCCCAACCGTAACCGCCCATGACCTCGGT
...
310446 TGGCTTAAAAATGAATACACAGGGGGAAGACGAATCTTTACCGGACTGGCATGTTAAAT
310496
310506 GACCTGGTTCTGTCTGTGCGGCAATGGGACTGAATTTCTTTGACGAATACGCAGAA
310566 GTGCAGAAACTACCGGTGATCGACGTGGCGTTGCAGCCACCCTGGTGGAAATGATGATGT
310592
310626 CAGGAACACTCACACTGCTGG
...
310753 TGAACTCGAAAGCCGATGAGAGAGTTCGTAAGAGCTTAATGCCATTGCTGACGTCACAG
310803
310813 CCGGTGCTGTTTGTGCTGTCATTAATTAATTCGATGAAAAAGATCCCAATGCTGATGCTG
310816
310873 CGGATGCCGTACGCCAGAAAGGTGCACAGCAATGCTGAACAGCGATGTGCTGCCGCCCTCCC
310911
310933 GCGTTATCCCGGCTCATCAGCTCAGCGCTACCT
...
311023 CCTGGGTGATGATTTTATCAGGGAAGCTTTCGACCTATGTGGAAGAGAGAGCTGGA
311073
311083 AAGACGACACTAAGTAATTAATAAAGCAGAAAAAACTCTGATATAT
...
311398 AAACCGGGGAATTAATGACGAAGGTCTGTGCTCACTGGAGGAGAGCTGAACGACTAAT
311448
311458 TGAAGAGCGGTAAACGCAAAAGAACAGCAAAATGCTGAGGAGAGAAAA
...
311671 CCAAGCTGCACCGTGAAGCTGAAGCACAAGTCCGGCTTGAGCTGGAGAAAAATGCTAGCG
311721
311731 GGATGAAACAGGTTTTTGTGCGACGGCCCTGTACGCCACAGAAAA
...
311873 AAAAATAAATCTGATGCGCTTATGAGAGAAAGACGAGACCGTCAACCGGTGCCGTG
311923
311933 GCAGAGTGGCGTATGGGGAACCGTGTGTCGATGTTGGCACCAGT
...
312194 GAGCTCTGAAGATCATTAACGCGATCAGCAGGCGCAGAAAGGCTTACCGGGGACTTCA
312244
312254 GGCATTAAATGAAGGGTTCTGAGCTCATTTACTGACAGTAAGCGCTGAGGAGAGACT
312296
312314 GGAGCAATGCTGTGACTTCACCATTTATTCAACAAAT
...
312424 CGATGTTGACACAGAGCATATCCGACGGCCGAGAGACGACGCTTCGCGGTTTAAA
312474
312484 AGAGGGGCTGACCGCGGAAGGGCGCTGCACGGCAGAAATGCCATCATGCGTTCGCTGCACA
312503
312544 GGGCGTTGAGGGGCTCCCTCGCGCTGCGCTGACGAGGATGACAGGCTGCTGCCACCAAAA
312545 312553
312604 CTACGCGAATTACCCGGGTAAACGACAGGGTCAAGTTCAGTGTCAATGATACAGCCAGGCT
312623
312664 GCGCATTAATTCCGC
...
312762 ACCAGGTACGGGGCGCAGTTACATCTTGACAGATATTTCAATGATGAGAGCGCGGC
312812 312819
312822 GATATGCTCATTTTTCCTCGCGGACGTTGTTATGTGACACTGGAAGAGCTTCGCTGGG
312824
312882 GCGCGTTTGGTGAATGATA
...
313129 GCAAGCTGCAGAACTTAACGGTTGCAATGCTGCTGCTCTGCTGCAAGAGCTTC
313179
313189 TGAAGTATGTCGATGAATAATATTCAGAAGGGGCTGCATGATACAGAAATGAAGACG
313249 GGGCTGTGCTGTGCGGTGGCGGGGGGGGGGCGCTTATTCAGGGGAGTGTGAAGCA
313279
313309 CTGGCAGAGTGTGTACACAGATCAG
...
313372 GCCATTATGCGGCGCTTCGCCGATGTCGAGAAGCTCCGTACAGCCTGAGAGCGGCTCTGG
313422
```

306944 GGGTACCGGCTTTTATTACCCCTCTGTAGGAAAAAGCTGATGAACGATCTGAACA
306989
307004 CCAGCTACAGGCTGTATGGAATCACATTACGGGCAACCCCTG
...
307124 CAGTCCCGGCACTGGCTGCTGACAAAGTTGTACAGGGGGAGAAAAACCGTGAACGATGAA
307174
307184 CACTGACAAATCATGACAAACAGATTGTCTTGGTACGGCCAAAGG
...
307290 ATGATACAGAAATGGCGGGATAGCCGGAACAACACTGTACCAACAATGTGTCAGGT
307340
307350 CGTGTGGAGGGGGGACAGCCAGTATAGGTTATTCGTGACGGC
...
307463 . ATGAGGGGGGGGTTGGCACCAGGTACATTTACACCGGCGATTACGACCGTTAAAA
307523 GTGGCGGCTGGCAACAGGAACCATATCACACCGGCGCAGAGG
...
307781 CGCTGAACACGGTAATTAAAGAGGGGGCTGGCAGGTTGTTAAGCGAGTGGCGCTGCCG
307831
307841 GTACACCAACCAATAATCAGAACGGTGAACCTGAGGGTACATGCCG
...
307977 TTTCAAGGTGAAAAAGGTAAAGCTGACGGTGTGTCTGAAATCCGGCGGTGCTGTGA
308027
308037 TGTACTGAGAGCCATTGACGACAGAAATACCTAGTGAATGACGGCGGTACCTGGCACT
308097 GTCTCCGGCGGTAAAGGACAAAGTGTACCATTAACATCCGGTGGTCCCTGATTGCAGA
308112
308157 CAGTGTGCCA
...
308266 AATGCCGGGGGACAGGCTGGCAACACACACTGTGGAATCGTGAACACTGACGCTGGCT
308316
308326 GCCGGGGAGTCTGAGTGGCAGAAACAGCTCACTAAAGGCGCATGTGTAATG
308365
308386 GGTGATGTGGTCAGTACCGGCGATATTGTTAAAGCAGGGAGATTCCCTTGATTAATGAG
308404
308446 ACGAACCGGAATGCCGCGTGAAGCGGTGCTGTTGCAAAAAGTAAGTCCCGGTAAAGTTC
308450
308506 CATAAAGTACCAACAGAACCTACCGGCCAGGGCGGCACCATCAATA
...
308445
308499
308890

308663 TCGGGGTGGCAACCAACCGACAGAGGTATCCGGTTGTGATGACAGAAATGGCGCACCA
308713
308723 CAGAAAGAGTGGCTTTGCCCTGATGTCGCCCGCTTACGCCCGCTTTAACTACACCC
308747
308783 TGAACCGTACAGCGCATGGAAGACTGTACTGGCGCAGTGAANAATGCTTATCGTGTGAAG
308794
308843 TCCCCCT
...
308891 TTCTGGCAGGCTCCCGCAGCATACAGACCGGTGTAAACGGTGAATAACACGCTCCGTC
308941
308951 TCAGCATTCAGGGGGGTATCTCGGTACAGATAAACAACGGCGGTAT
...
309040 GTCCGCTTGAGAGGTGACCTGCTCAGAAACAGAGTTGCCGTATGTCCTGTGACAGAGG
309090
309100 GTGTATGTTGCTGACAGGCATTCTTCCGTTGATGTTAAGATGATG
...
309220 TCCGGCTGTGGGCTGACATTTGGGCCAGGAACCCGTACAGCATGAAGGCTATCG
309270
309280 GACAAATACGACTTCCGGCCCGGGGCTGGGGCTGGCTGCTCACTGAAAAACGGTCTG
309286
309340 CC
...
309374 CACAATCGAGTACACCTGGCAGGACCTCTCCCTGATGACGGCCAGATTAACGCGGTT
309424
309434 ATGTGAAGTTGGGCAATGGCAGTGCACAACATGTGCTGCCGTTTCCGTCGTGGCAGCC
309494
309496
...
309616 AGCTCCGGGGTGACATGAGCATGGGGACAGCCGACGCGCAATTAACATGACGTTCA
309666
309676 CCGTCCCGGAATGCAAGTCACTGAGCCTGCAGCGCGGCACTGGAAGCCGTAATCCGGGAA
309736
309748
309796 GGCTATAACGGTCAAGCTAACGCTGAATATGACTTCTGATTAATACGTAATCCGGGAAATA
309872
309856 CGATTTCAGAAAAATTATGACAGTAATGGAAGAGCGCTATTGCTGTTCACCGCAGTT
309890
309890

304118 GGCATAGCAACATCCAAAGTCTTTCCAG
...
304168 CCCCCGTTTCCGTGAGAGTGTATATCTGTGMACTCAGGCTGCCAGATCATCGTTTCCG
304228 ATGSAAGCATATAAGCTTTTCTGTCTCTGCGGAGAGATGCGCCAGCCTTCCAGC
304288 |-----|
AATCGTGATTTGTATPACCACTGCACCCAGAGTGTGGCCAGTTCCACTTTCGA
304289
...
304423 AGCGTAAACTGACCGCGTATGTAGCCATCAGACGAAAAATGTAACTTAGACGCCCATC
304483 TGATATAGACGACATCTAAGTATGAAATTAACAGACTGGCGAAAA
...
304639 CCTCCTGTTCAATAGCTCAGGCTCTGGGAGAACGAGGCGTATATCAGCGGCTTCA
304699 |-----| |-----| |-----|
GGTAAGGACCTTCTCTGACACTGGCGGTGAATTAATACCTGTAGAATAACGCCGATGA
304702 304721 304747
304759 GCAGAAAGAACCGGTGGCGCATTTGCCCGCTTTATCCACTT
...
304811 AACAGTGTACGTGCCGCTCCTGCAAGGTGGAGTTCCTGCAAGTATGACATGACGCTTGA
304861
304871 |-----| |-----| |-----|
AATCCATGCGCAGAGCTGCTCAGCGTGTGATCCGTGAACGTGACCGGAGGAGGACGATGA
304876 304923
304931 TCTCACTCCCGTCAAGGTACCCGCACTGTGCTGTGGCAGCTTACCGATATGCGTAAT
304969
304991 CCTTCACGATACGACATATGCTGAAGAGGAG
...
305090 TAAATCTATTTTGATGAAAAATTATTAATGTTTCCAGATTTTAATCAGCGCCGAG
305140
305150 TTAACCAAGTATTGAGCGCCGATTAATGTTTATTCATTAAATCTGCATGATTGTCT
305165
305210 TATTATCAGCT
...
305275 CTCATCATTAATCTCAGTATTATCATACGCAAAACTTCTCCGAGACAGGCGCTTCAAG
305325
305335 |-----| |-----| |-----|
ATAATAGTCAATCGTCAATAGTTTTCATCGTCAAGATAATGATATGTCGATTTCTGAATATGA
305344 305359

305395 TTCACCACAATTATACATGA
...
305419 ATCTCCTTTGGGGGGCAATAATTAGTACGTTTATTCGAAAAAATGATGAGCGCCAGAC
305469
305479 CAATCCACCAACTTTTGTATTAATCATGAAGATACCTTGTCTAG
...
305623 GTATGATTATCAGAGAAATACCTGCATCCTAAATAATTTACAGACAGATATTAATCGTTGTA
305673
305683 ATCCCCCGCACCTTCCGTGCGGGGTTTGTCTTTTTCACC
...
305914 GGGGAGGTCAACCCGGTCAAGTATTCACGCCCGGCAACCGGAAAGTGCGGCGCTTCCG
305964
305974 TCTGAGTGGCCATGGTCAACAGCATGTTATCAGTACCTGCCCCGGG
...
306097 GTGCGTATTAAGCCGATGACCGGTGCCCTGTCTGTGATGATGATTTCTGGGACACAT
306147
306157 ACTGCACCGGGACATCAAGAGGTGCTGTTGTGTGTAACGACGCC
...
306335 TCGGCCCGCACCGGCTGGGAACGTGATACGCTGTCTGTGATGATGATTTCTGGGACACAT
306385
306395 GACATGCGCGCAGTCCCTGATGACCCGCTGCAAGACGAGCTGT
...
306465 CCGTGAAACAGTTTACCGGTGCGGTGACCGGATATTGACACGCTGGAGACGCTGTG
306515
306525 TTGCTCTGTGTGACACGACAGTTCTGCGTGGCGGTCCGTGACAGGCT
...
306680 CATGCCCTCCTGTTTCTTGTGCGTGAAGTCTGCTACTGTATAAGAAATAAAGCATGAT
306730
306740 AAAACAGTATCGATCGTTTATATGATCGTTTATAGTGTAAATGCGCTGGAAAGAGC
306753 306766
306800 AGCTGAATCTCTTCATCATGA
...
306824 ACGGAATTGCACACAAACAGACTGATTAATCTCTGTGTGATGAGAGGAGGAGCGTTTACG
306874
306884 CCGGGAGCTATTCGCTGTGCCGATTAACCTGTTTTTGTACCTGCCGGTATCCACTTTTGT
306894


```
300059 CCATTGCTAGATTACGTACTATAGACAGTGTATAGAGGAACTGGGACGCTTTCG |-----|
300109
300119 TCAGGCTGACAGGCTAAATCTGAATAAAAAATATTCAGCGGGA
...
300647 GCGTTTAAACGAAAAAATCCTTACCAGGAAATGTTTGGCTGTCAATCGCCCGCA |-----|
300697
300707 TTGAGTGAATTTTGAACATTTCCGTGATGTCTGCTTTTC
...
300762 GGATTCCACCGTGTTCATCTGTGCGAGAGGTGGCCCGCAGAGAAACGTCATTT |-----|
300812
300822 CTCGTTCTGTTCATTGACTGGAGGCCAGATCGGTGCACCAAT
...
300902 ATGTGACGGAACCTTTTACTGGGTGCACCTCCCTGACCAAGGTAACGGCTCTGCG |-----|
300952
300962 AGTTTCAGCCGGAGTGGATATGCTGGAGCCAGAGTGACCTGGGT
...
301021 GAAGAGGCCATTAGGATATGACATATTTCCCTTTTCGGTATGCCATGACGTTTGA |-----|
301071
301081 GAATTTGTTCCGGCATTTTCTTCTGTTTCCCGGTAAACGGGTGGAGTGGAGTGTG |-----|
301141 ACCGGTGTTCGGGATGAATCCCTCAATCGCTTTATGGGGTGTGTCTCAGCGCAA |-----|
301169
301201 CTGAGATATGCCGATGATAGCCCTGACACAGCGCATCGCTGAAGGTTTATATACCC |-----|
301236
301261 ATGTATCCGTTGTATGACTGGAAGCCCGCATATCTGTGATATATAAGCCAGACTTGT |-----|
301306
301321 GCCATTTACAATCCTGTGTATGACCTGATGTACCGTGCCGA
...
301430 ACCATGTTCTGAGACCGGAACCTGGGCCAGAGATGTGAGCGGGTGTACGGCCTGCA |-----|
301480
301490 GCGGGGCGCTTATGCAATGAAGCGGTGCCATTTTGGCCCTGCTAAACGTATATCA |-----|
301494
...
301767 TTCTGTGCGCGCCCTCTCTCTCAGTCCGACACAACCCCGGCACTATGACGTTATCT |-----|
301817
301827 GCACGCTATGAAGAAGAGGAAGAAATGGGGATTTCTGTGACACCGGAGATAATTGGC
301887 AACATGATTCGTAGTACTTTCAGCCAGGAAGCAGACAGAAAGAGACCATACAGCCCTT |-----|
```

```
301947 AATCACCCTGCCAGAGTGCCTTCAT 301915
...
302013 TGGGTAAAAAGCATGAGGTGGTGGCCAATGATTAACAATCTTAATGTATGGCCGGGT |-----|
302063
302073 GAGAGAAACTGTGAGCAGTCACTGGAGAAAGATGTTTCACAC
...
302123 GGTGTGTGTCTGAGAAAAAGAACCATTAACCTGTGTGTGATGTTTTCATCGCCAGGT |-----|
302173
302183 GCTGGGAGAGGAGAGGTACGGGAAMACGTCGTGAAGGGCTGTGCCGTGTCTGAT |-----|
302209
302243 TAATCCGGAGAGAAAGGAGACAG
...
302661 GCACAACGGAGTGTGTGTTGTTGCCCGGAATTAATGTCTGGCCGACATATCGTCCCG |-----|
302711
302721 CCGTACCTGAACGCTGGATATCAGCGAGTGGCGAAAGCTGCAGATGCCAGCTGAGG |-----|
302732
302781 CGAATGG
...
303550 TAGATGAAGTGAATGTGTACTATCATCAAAATATGAAGATGGCGAGGAGCTTCAG |-----|
303600
303610 TTATATAGATATAAACGGTAGTTCTGTGTGACCCCAAGAGTATCTGGCAGGGGAAATG |-----|
303670 ATTGATGCCAGGGCCCTAAGCTGTGTGTGCAAGATTAATGCTATAGACATAGTAAT |-----|
303674 303687 303690
303730 TTTTATGATTAAGAT
...
303818 ATTATGAACAGGAATTTACAGCGGGCTACAGTGGGTAAACCTCTATGAACGTCGGT |-----|
303868
303878 GATCGGGATTTGTATGCGCTGTGTGTCATTTCTGGCCACATTCGTAATAATGTCG |-----|
303897
303938 GCGCAGATATCTGGCTCAGGCGCATGCTGTGTGAAAGCCACAGCCGTCGCCCAAGCG |-----|
303960 303985 303994
303998 TTCAACATCAAGAAACCGGCACTTGCGAAGTCTGTGATTCGTGAGTTACGTTCTCA |-----|
303999 304029
304058 GCGAATCTGGGGCGAAGACTATTCAAGCGAGTTAAAGCACTTCAATATCTCACT |-----|
304096
```

295735 CGGCGAGCTCAGCATTTTCAAGTTCAGACTCTTTCAAGACGCTGACGTTCAAGCGGTGCTG
295772 295778
295795 AGTCACCATCACCACACTCCCGGTATCCGCTCATGCTGA
...
296087 TGGAAATTTATGTTGACAGAACGAGAGACTGAGAACAAAGCTTATTCGCGCTGTCC
296137
296147 GTCCCTTGCTGCTACAGACAGTATGCTGCTTTATATGCGCTTATCTCTTCAGGACAGC
296183
296207 GGATGTCCGGCGTAGCGGAGATGAAGCATTTATCATTCAGCAGCAGCGCTCAGGAAGCCCT
296251
296267 TGAGCAACAACACTGAGCGCTTCAGCCCTGATGTGCTTCTGACACTGGCTCTTTGC
296279
296327 CCATAGATTAATTTTCCCTGTGAAGCGCCCTGTTTCAGATTAACAGACGGAACCTGA
296351
296387 GGGGGCTGATGCGTTACCACACTGGCTGCTTTACAAAATAATCCCAACGGGGGCGTGG
296427
296447 GCATTGCCCTGGGGGGAAGAATTAACTGCTGAT
...
297459 GTCAAGATTTGTGTTGACGAGTGAAAAATTAAGACTATAAAATTTACGAGCTGAGCGTGG
297509
297519 GCTGAAATGACGACAGAGAGGTTATGCTGACCTGTCTATTTTCCAGCGGGAGAAAAATAT
297528
297579 ACTGCTCCGCTTCTGTATGTCAGATTAACGCCAAGTGAATTTGCCAAATGTTTCTCTC
297605
297639 GAAAGAAAGGAGACGCTGTAACTCCCGGCTGAAGTGTATTATATGATGATGAATAA
297651
297699 TATGCTGCTGTTTATTAACCGGAGGAGTGAAAGCACTTACATCATAAGAAACGACGCGCTT
297727 297746
297759 GACTTTCACAGGTGTTACAGATTTAACAGAGCATGCCGTGTTTA
...
298193 TATGTGATGGCAGACTGGGAGATACGAGAACAGCAACAGTTTTCCTCAAGCTGCATTT
298243
298253 TCGCGTTCTGAGGAAGTCTTACTGACTCACCGTCTGCAAAATGCGA
298299 TTGATGATGTACTTCAACAAGTGAAAAAACACCTGAATGCCATTAAACAGGTGCGCTTGT
298349
298359 TAAAGCAGAAATTAATGAGCTGAATATGAGTATAAATACCTGCA

...
298553 AACTTCATGATACGAAAGTCAAACCATTTGATGATCCGGAGAGAGAGGCTGCGCTATTT
298603
298613 ACTTCTCTTACGCTTCTTATTATGAAAAAAATATCCGGAAAAAC
...
298669 GAATGAATATGAGAAAGCATTTAAAAATCTGATTTTAAAAATTTTACTGGGCGCGCTGAC
298719
298729 ATCATGCTCTATGCTTTTATTAAAGCAACATCTGCGAAGAAATGTT
...
298783 AAGTTCAATTCGCGCAGACATACAGGATGAATTTGACCGTTTATATAAGCGTTTCC
298833
298843 GATTATTGCTAGCAGTACGCATTGCAATTATTAACTATATGCGCAAG
...
298929 AGCAGATATTTGTTCCGGGCATTTCTGGGACTGGGATGCGCCGTAAATGATATGCTGTAG
298979
298989 GGGATCGTAAAGCACTTCCACATGTTCTGTATTACTCCGAGACAG
...
299092 TTGTATGTTGTTCAAGAAATGTCAGCATGTCAGTAACCTTCTTAAAGACATTATGCTGCA
299142
299152 TCCATAAATCATTCATTTTGTATAATAAGCAATTTTATGATTAATGCT
...
299484 GCCAGCAGTCGGAACATCGCTTTGTGATTAATCCTGAACCTGTAATGCTGCTGTTT
299534
299544 CAAGGGCCCGAAATTAATTAAGCTGTGACTGTAATGATGATTAATTTGAACGCCACGCTG
299592
299604 GACATATTTGGCGCGCTGATTTCGTACATCAAAATCTACGCTGATGACGGAGAGATTTTGC
299614
299664 AGAGCC
...
299671 GGTCAATATCAGCAATTTGATTTGCTGTACAGTGAATGACAAATCACTGGAGACGTGA
299721
299731 TAGTAGATTTGAATTCAAATGATTCACACTTTAACTCAGAAACAGATT
...
299795 GATATCCTGTGCGCAGATTAATTCGAGATGATGATTTTCATTTCAAAATGACGCTGAAT
299845
299855 CAGTTGCTGTTGCTGAGCGGGGGATTTTACACACCGTGAACAGT
...

```
291982 AAAAGCTTATAATGACTACTCTTCCTTTCATTTCTGAAATTTCAGC
...
292380 ACATTTCTGCAAAATACCGAAGCCTTAATTTACAGCCGGGGAATCCGGCGTTTGT
|-----|
292440 TCAGCAGCGGTGACACACTGCGCAAAATCTGTGGCATTTTCACCTTC
|-----|
...
292549 CGACGGGTACACGAACGTAATTAAGGACATTAATCACACGGTACGCCCATCAGCGCGGTG
|-----|
292609 GTATTACGCTGCTGCGGGAAACATGAAAGAGCACAAGATACATGCTGTGTGCTGCTC
|-----|
292669 AGCCAGACGAGTATTTCCTCATCCGCGAGTGGCGTCATGCTGATGAAGAAATGTGCGCA
|-----|
292729 CTGCTGATTTACTGCAACAGACCTG
|-----|
...
292786 GTGAAGAATGTTTTGCGGTTGCCCCGCGCTGTTTCATCAGGGCAATACGTGCGTTAAAA
|-----|
292846 ACGGCTGTTCAAGTCCTTATGTGCGAAGTTATTGCAACAATCATCC
|-----|
...
293202 TTGCCGCGCTGCTGAAGATGGGCGATCAGCCAGAAAAATGCTGTTTAACGATAAAC
|-----|
293262 GGTTTGGCTGGGTACAGCCTGAAGGCCAGCCTTTCTGTCGTGCC
|-----|
...
293354 GGGAAATGA AAAACATCATTTGCCATATCAGCAGGTTTGGCCGGGAAGTGAACCGCCCGT
|-----|
293414 GAACGCTGAAGTATGCACTGGAAGCGGGTTTACTGATGACTGCAC
|-----|
...
293754 GAATGAGTGGCTGCGAGAGAGAGATTATGCCGACTGGTGATTAACCGGAATCCGCTGGC
|-----|
293814 AGAAGTTGTTAACCCGCAAAATTGACAGACAGCTTTTATTAACGTTTCCAGGGATGCGCT
|-----|
293874 GAACACGCGCATACAGTGAACACCGGAGTGGTTTC
|-----|
...
294081 AAACACACCCGGGTTTCTCTCAACACTGCTCAGGGTGACAGCGCACCATGATGCTTTG
|-----|
294141 TATCGCTGATTGACTTTTCGTAATGGAACATCTTCTGACATAAATTTTACTACGSTA
|-----|
294142
|-----|
...
294319 CCTCTTCATGTCCTTCCCATATCCGTGAACCTCACTTAATAATGCTTGAATCGTCAAAC
|-----|
294379 ATTGCACCTTTCATGACTGAATCAAAAATTTGCTATTTTAAAC
|-----|
...
294667 GTAATAACCGGAAGGGCGGTAGCGTGCATTCGCAAAACATGGTGGAGGATACGTCAGGA
|-----|
294727 GGCTGCTGAACCCGCCCGGTTTCTCTGAGAGTGTATTATCTGTAACCTCAGGCTGCCAG
|-----|
294787 ATCATCGTTTCTGATGGAAGCATAAATAAGCTTTTCTGCTCTGCCGAGGATATGACC
|-----|
294790
|-----|
294847 CAGCCTTCCAGCAATCGTCGATTTGTTATACGAGTCCACCCAGCTGAGTGTGCGCAGTTC
|-----|
294907 CACTTCTGCA
|-----|
...
294991 ATACGAGTCACCTGTACTCCCTGTTGATGCGCAGCAGTTTGTCTCTTTTATGCGCTCGT
|-----|
295051 ATAGCCAGTGAATATATCTGAAACCTTTATCTACTGTGATGAGCTGTGCCGAGCGCG
|-----|
295111 ACGGCCACACAGCGCTGTGCCATGTCATCCAGCAGAAATGCTGTTTCCATGACGATGA
|-----|
295121
|-----|
295171 GACTGCGACCCCGAGATACATCCGCAAAACATCAATGATGAAGCCACATAGACGAA
|-----|
295231 GCCCTGCCATGTGCTGACGTAAGTAATAATCAGCCACCCACAGCTGCTCAGGCGTTCTGC
|-----|
295291 CACGAACGTGACGTTTACCGCGGTGCGCTGCCGAAACGGGTTTCCGGGTGACGGTATGACG
|-----|
295351 GACCTTTTACC CGGA
|-----|
...
295396 GAGCGCGCCACTGTACATCTGGCCACCCCTGATACCTTTCGCGTAAACAACTGGCGCCAGAC
|-----|
295456 TTTACGACACCGGTACACCTGATGATTTTTCATGATGATAGCGCGCTGT
|-----|
...
295510 CTTACGCCAGTCAATCGCGCTGAGCAGGGGCACTGCGTTTATCAGAGATGATGCTGCTTGG
|-----|
295570 CTGACAGTGTATATACGTTGACGGGCAATATGACATTCTACTGCAT
|-----|
...
295675 AGCGGCTGAACCTCGCCTTCGCAAAATACGGGAAGCCTGGCGAAGATATCGTACTGC
|-----|
295725
|-----|
```

```
288645 GTTACTTTTACGACGGGGTTATCCAAATACCTTTCTAGTCAGTGTTAGTATTATGCAA |-----|
288647 288704
288705 GCGATATATAGTAGTCACCGCTTAGACTATTGACATTCATGATGCCGTTAATTCA
288709
...
289051 TTTCTGTGCTCAGCGCCCTTTCGTGTAATTGACATGATGTAGTAATTAGTACCGCTGCA |-----|
289101
289111 AATTGCGAATTAAGCAGCAGAGCTGAGCTGTTGTTTGTATGACAGTCCAGGTTAG
289171 ATTCAAGAGCTGATATCGCCATCAGAGAGTTGAGGAAAAAGTAATATTATTGTC
289186
289231 ATTTCTCTTTTTTAGCGATGAATGSCAGATTAACCTCATGATTTTGCAGGAAAAATGA |-----|
289245
289291 GGCCAGATTT
...
289559 TAACGAGAGTCTCTTAAGGAACATGCTGGACAGGATTCGTGCGGAACGTTTACT |-----|
289609
289619 TATCAGCCTTGCACTGTAAACCAATGTTGATGTCTGTGAGCTG
...
289803 AATGAGCTGATGATTTATTCATCAGGTGGCCAGGTAAATACCAGCGCTGAT |-----|
289853
289863 AGTGAATACCACAGATTGCTGTTGGCGAAGGTGCAGATCGCTCTGTTAATGCAAAAT |-----|
289900
289923 GTTCGAAGCGTGGGTGGGATTAATCCGACAAGCATGCACTGACCAGTGAATTTGTA
289983 AATGCCCTTCTTTATGATGATTCACTACGCTTCTGTTCAATCAAGTAGCCGCTGAA
289984
...
290192 CTAAATTAATGAGTAGTCTCCCATTAATATATTGCAAGAAAGCTTATCTGACGATAAA |-----|
290242 290250
290252 CGATATTTTACACAGAGTATATTATTGATACCTTCTGCAATTATTAATTGCTA
290255
...
290355 ATAGACAGTAGGTTACTTTCCAAACGACGCGATTAAAGCAATGATGCAATCGCTGAC |-----|
290405
290415 ATGTGAGACGGTTGTACCTTCTAAGAACCGACTCCAAGTCGAATATTTCATCTTCGG
290475 AGCGCGTAAGAAGTATCTGCTACCTTACCAGAGAGTATATTCTCAGGTAATGACTC
290476
...
290553 TGTTCAGATTAATGCCGATGACCTTGTGATGAGCTCCACGATTTTGAGACGACGTG |-----|
290603
290613 ACTTCGTCACGAGCCCTTCCAGANGTGTCTCAGATTGATGATTAATGATGCGCT |-----|
290658
290673 GAGTGTAAAGCTTGTGATTAACGTGCAGCTTTCCCTTCAGGCGTATTCATACAGCGCC |-----|
290679
290733 AGCCATCCGTCATCCATFACACAGACCTCAAAAGCCGACAGAGCTCAGAAAGAGCTCA |-----|
290784
290793 GTGTGGCCAGAGTGCCTTCACCGAAGAGCTGGCCACAACCGCTCCGTATCTGTGAT
290853 ACGCGTAACAGACGCGCTGACGTGATTTAGCACGACGAGGCCACTGTTGTGCA |-----|
290868
290913 TTTCAAGCGCAG
...
290959 TACGACTGCGGCTCGAGTTTTTTAAGTGACGTAAACCCGTTGTGGCCACAGCCCTA |-----|
291009
291019 ATGGGTGACATGGCGCGACATCCGACGCTTATCATGGCCATATCAATGATTTTCTGTGC |-----|
291042
291079 TTACCGGCTGAGAGCGGCTGTAACTGATGATGTCATGTTTACGGCAAGAGAG
291139 CAGAGATGAGGCTGATGTGCGGCAAGTCTTTTGGCGTTAGCCACACGCGCTTCAGTAGCG
291146
291199 GAG
...
291338 TTATAGTTGCATTTTATTTTGTGTTATTTTCATTTATATCCCTTACTTCTGTTAA |-----|
291388
291398 ATTCTGTATATTAACAGSAGAGGTAAACCATAGATGAACGTTTATTTTCTGTATCAGA |-----|
291435
291458 ACAGCCATTAACCTTTTATTTTAACTGTATCACTGAATGAATGCAACCTAATAGACAA
291518 TGAGGAAACAAATGAGAAAGGAGGAGCGGACGCGCATATGTTCATATTTCTGTGAT |-----|
291539 291551
291578 CAGCCTCAGTATCAACCTTCTGAATG
...
291748 GGTAAAGAAAGTCATATACTCTGTGATTAATCATCATCGATTAATAGAGCGCAGCA |-----|
291798
291808 CCTATCCAGTATGATATGAGAAGATGCTCATTAATATGCGTCAG
...
291922 ATTTTCATATAGTAACATATATAAAGTGTCTACTTTACTATGATGATGAGCATTTAT |-----|
291972
```

```
285161 ATAAAAATATCTTTCTACTACGAAATATTTTGTGAAGTATGTGACACGCTAAGT
285221 ATTTTAAATATATATCATGATGCTTAGGCATACATTTTCATGAGGCTATTTCCCT
285281 CCGCGGACATATGTGTGCTGCGCTTTTATCATTTCTACTTTTAAATCCACTGT
285341 CCTTCCACAGAAACA
285865 CAGTATGTGACACTAATAGAAATAAAAATCTGCCGCTGGCAATATTTAAACGATGCCA
285925 TACCGTATAGTGTAGTCAAAAGTAACGGGGTATTTAAACCATCTT
285981 TTTAACAATATGCAAGATTCTGATGTATTAGTTCAGTTGTTGCCAGAAATCAGCGCTGAT
286041 CACTATGGCTTTTAGCAGTGTGCTGAGAGAAGTTCTTCACTACTGC
286141 AACTTGGCTATAGCCTGCAATTCGATATTTGCCAACCCGATTCATATGACAGCGGCG
286201 CATCCTGTGCGCGGCGACTGTGATGCATGATAGCATGATATGATTAAGCATGCTCCATCA
286261 CCGACCTTTCTGACAAGCTGTAGAGATTAATTTCCCGGTTAAACAGATCGAGAACTGAC
286321 --| GACAGTACACACTTTTCCCTGTACCGAAGACTGTATACATCCGCTCCATTT
286434 GGACACCACTTTGTGCGATGCTTCAGATTATGACAGATGATTTCCGCTGGACGCTTACG
286494 ATTACATGCAACCATATGTGTCAACGATTTAGCCGACTGATGTATGACACATTAAGGCT
286554 | TAGGCTCAGGCCCTTATTAACAATGTCACTAGTGTGCTTGTAGCGCGTAAA
286622 AATGACCGGCAAACTGCTCAGTGTATTAAGAAATCAAGTGCCTGATTAACGATGCCA
286682 CATCTGTGTTGTTGGAATATGACAGATTGTTGTTATTTGCTTCA
286760 GATTTTTTGGCATTATTTTAAATGAGGCTTCAAGTGGAAAAATAGTTAGGCGCTTCAAG
286820 TGGCAGCGCGCAGCTGCCCAATTTAGTGGCAAACTCCCGTTGGCGT
286947 AATGAAAAATGTAAATCAGTGTGTGAGCTGTTTATCCGCATTTGGAATTTGCTCAACT
287007 AAAGAGATGAGGCCCGTTTGTATCGCAATGGGTCTATGCGGCTCATGCGGCTTGTCT
287067 TCGATATAGCTTCTCTGAGCATACGGTGTGCA
287258 GCCTGATTTAAGCCCGCGGTGGTATTTGTCCAGAATGCTTCAAGCTCCAACTGTTTGG
287318 CCTGTGAATGAGATGGGTTAAGCAAGACTCTGTTGCTGTACTTTTTCGCGCTG
287378 GCGCTCAAGCATGATTAATGCGATTGCTGTAGTTCTTGTGATGTCATTGTCTGC
287689 GACGATGTGCTACCTTCAAGACCTGCTACCTATACGCCGGAATTAATGTCAGAGAGCAT
287749 ATTTGACGCCCCCTAATCTGAGGCGTTTGATTCAGAGATGCTGGAACAATCCGGGCTTTG
287809 TGTCAACGGGGCATCAAAAGCTGTACACCCGGGATGTGGCGCTGATGCTATTACCG
287869 GACTCGAAGACATTACTTAAGAGGTGATATGGGGTTTCAAGTATTTTGTCCGCGATAAT
287929 CGTGTGTGTATGATCCGGAAGCCGTT
288311 CAGCATGAACGTGTTGTAATGCGGTACGCTGGCAATGTGCAAAATGCCAGCGCTGTA
288371 TTGTGAACCGGCGCTGCGCTTGTGTGCTGGGATATTTCCGTTTCACTTCTGCTGATA
288431 CCTTCCGTTCAAGCAGCGGTTGTTA
288525 TTTCTGGGTGTCTGCGCGGTTTCAGTAATTAAGAAAGAAATTTGTCAAGCGCTAGTG
288585 ATATAAGCGTAACGCTTAACAATCTTGGCGTTAATACATTCCGAGCAATAACCGAAGC
--|
```

```
281023 CTGACAAAGTATGTAATGATTAAGTTAAAAATATTCGGTGCATTTCTGTGACGATATTA 281073
281083 GTTGTGGTAACGAATATCTCATTAGCGGATCGGAATGCATTAAAC
...
281140 CCACGTGTTCCGCCAGGAGTTGCAGTGGCAACGATATTCACGCTCAGTGCATGATACGGTTTGT 281190
281200 TATTATCAGAACTGATAGTTTCAGCGACATCCAGTAATTTTGGCTCAGGTATTCACATATCTCT 281221
281260 GTTGTGATGATCCCGT
...
281324 GGCTCACGTCACGTGTAATCCTTGTAGCAGGTTATTCACGCGCATTCAGCATTCAGCGTACAGG 281374
281384 CATCCGTCATCTTCCACGACAGACCTGCATCCACCAACGCTCCAGGCTTTCAGGTATTCCTC
281444 CTTTCTCATCATACACTTTCTTCTGTACAGCGCTCAGTGACGATATTAATTTCTGGTGAAAC 281472
281504 GTGGTGTTCCTCCGCGATACGGG
...
281818 TTGGCATTTTAAAGCCGCGGGGTTATCGTAATACACCCACGCTCATTAAGCTCTCTCT 281868
281878 TCCGGCTTCAGGTCGGGGTTACCAAGTAGATTTGTTTTCCCTGCC
...
281948 GTAGCTGCCCAATTCCTGGGTGCTTATATCCCGTGTGCACACCGCTTTCACGCTCCAGG 281998
282008 CATCTGCCACATCCCAAGACAGATATGCAGCGGAGCTGAAGTGTCC
...
282462 CCGGTCAATATCCCCCGTCAGACTCCCAAGTTGCCCATCCGGTTATCATACGCTGTGCCG 282512
282522 GGTGGTATCATATCAAAACAGAGCACATCTGTCCGACGCGTTCCAGTCAAGACGTGC 282559
282582 ACCAAGATTATTAATTTCTGTACTCCGTGGGATAGGAATACGCGTGGCTGCTGTAATCGCT 282636
282642 CAGTGATGTGAACGATGAACCCCTGACGCTGTTGTGTGCTACCGGCTACCTGCAGGCTGAC 282665
282702 AGAATCATCCACAAGGGGA
...
283365 TATTTGCCTTTACAAAAATTAAGATTATTGACAAAAACAATATTAATTAACGCTGCC 283415
|-----|
283425 TCGGCATGCCAGTTAAACAGCTATCCGTATAGCGGTGTTCGTTTGC
...
283489 GTAGAAGCGGATTTCCACCACATGTCACCAATGAATGTCACTGATTCGAGACGACTGGG 283539
283549 ATGATTTTGGTCGATGACTTCTTCCGGCCAAATCTGATAGAACAATTTGAANAATTTGTAAAGA
283609 GAAACGTTCAAAATTTCTGTGGCAGGTGAAAAAATTTCTACAGCGCTATACCTTATTCACT 283612
283669 CGACCTCAAAATTAACGTTTAAACAGACATCAACCATTTGGATTGTCCGCTGAGGAG 283683
283729 CGAACAAATTTCTTAGACAGCTTTATCATGTCAAGCCACATGTTGTGGCGCTGCGCAATACTA 283773
283789 AGGACAAACGATGGGGAAGTAGCTTTTATAGCGTTATTTTCCGGAGTAAAAAAGTGTAAT 283795
283849 CTTTCCGTATGTCACTGTGTATCTTGTGCTGATTCGTAGTCGTATGAAGCATACACGAT 283903
283909 ATTTATCTTAGCTTAGTGAAGAAATAGTTGTTTGTGATCTATGTTAATCTGGGATTAC 283918
283969 AATCC
...
284155 CGCATACGATAGTGAATGAATTAACGAACAATAATTACGTCAACAACACTCAGCGCTTGC 284205
284215 CTGACATGCTCTCTGACATCATTTGATATTCACGCTCATCTGCAATATGCAATTTGCCCTCTT
284275 TATGCTGAATTTCTGCCAATGTTCAATTTACAGGCTCATACAGATCCGTTATACTCAT 284306
284335 GATAAATATCTAACATATATACCCGTGAG
...
284687 AGCATTAATTTGCTGGGAACGTGACATATAATTAATCTTCTGTGTAAAGGAACGTTCTGA 284737
284747 AATGTTTAAGTATGAGAGATGTTGATGGCGGAAGAAGATGATAA
...
285003 AGGAGTGAGATGCTGTGCGTGTGCTATTAATTTCCCAATTTATTTGGCGCTATTC 285053
285063 TTGCTGGCGTAGGCATCCGCATACTATGATAATTTGTGTGCTGATCAATTTTATTT 285077
285123 TATAACTCT
...
|-----|
```


275038 AAATTATGATTGAGAAAGTAGACTATGAGTGTCTAAGCTACATGT
...
275412 GTTGCGGATCCGGTTTAAGAAAAAGGCTCTTCCTTGGAGGATTGGTTTGGAGCGCAACACATCAA
275462
275472 TCGATCTCGATGCTGGCTGTGTATTAAATGACACAGCACAGGTAAACAATTACACCATTCT
275532 GGTTCGCCAAACTGAATAATCACCCTGC GGCGCTGTGGTGTGACACAGTGGGATTAACCTGACCG
275592 CTGAGGGTGAACGGTAGCATGAAGAATGATTAAGTCAACCTCTCCCGCCTGCCTGCACAAACG
275652 -| |-----| |-----| |-----| |-----|
TTGAATACCTGCGCTTACGGSTAACAACGCTTCGCTGCACTGCTTCAATGACCTGAAA
275653 275662 275692 275702 275711
275712 ACGCTTCTTCGCGCGGTTGTGATCAAAACCGCAAAAGACTGCGCTACAAAGCTGACTG
275772 AGCAGGGCTCACAATACCGGGATCGTATTTCTTGGCTGCGCCGCTAATACGGGCAACTGGC
275832 ACTTACCGCCCTTGGCCATGCTGTGCGCGCCGACCAATTGACGATATGCACTCCGATA
275892 TCCTTTCAGCGGTTATTTCGCTAATGAACCTGACGCGCGGGGGCAATGCCCGCTTCAGC
275952 CCAGGAGTTGGGGGTCCGGATCACTCGGGGGGCGACAGTGGATGCTTGGGGTTCCGCT
276012 GTATGCCAGCGGAAGGTGCAGGGTGATGGCGACATGCTGTTTATGATCAGCACGATA
276072 CGATGATGATCCGTCAGCCTGCTGCTAGTGAAGGCCAGTACTCAACCTTACTGT
276075
276436 CCGGTTGTAACCCCCTCCGCCAGTTGAGACAAAAATCAGCCTGAGCAAGGATACGCTGACA
276496 AAAGAGAACGCTGCAATCAGCCTTGACCAAAAGGGATTAATTTCGGTG
...
276625 GATCTGGGGGCTTTTGTGAGCTTCAGAGCGGGTATTAATTCGGTCAATTCAGGGCTCGGT
276685 AATGCAATTCCGTGATTATTCGCGATGAACCGTATGTTTCAGCTCAAGGCGGATGACCGGACG
276745 GGTGATGATATCAGACGGCGAGTGGCTGCATATCAATGAGACG

277004 TCAGAGCCGAGAGCAAAATGACCGCGGCAATTGGCTGGGATTTGCTGGAGCGCGTT |-----|
277054

277064 CTAAATTAACACAGCAACAAAGAAACAGGATATAGCTTTTCGACA |-----|
277346 TTTCGATTTGATGTGTGAATTTGGCAAGGCGAAACCATGAAATACATCCTGGCGCTGAAA |-----|
277396

277406 GATCAGCCTGAGGCGCGCTCAGCTGGCGCTTACGTTGGTATTTGCGG |-----|
277455 CGAAAGGTGACGGTAACTTGGATCAGAGAGAAACTGGCCTCCCGCGAGATGCTATTCG |-----|
277515 CGTTGGGCTTCGACCGGCTGAATTTTGGCCTGTGATCCATCCCT |-----|
277823 CGATTAACCTGTGTGATGATGAGCAATCTTCTCTGTTCCGCGCTTCGGATCGTTATTCG |-----|
277873

277883 CCACCGTGTCTCTACTGGGGGATCAWTGGTGCATGCTTTCAGAGGGGCATCTTTGTGCG |-----|
277938

277943 CATCGTACGAGCGCTGCTGAGTCTGGGCGGTAGCTTGAAGTTGTCTTCTGCTATTTATCGT |-----|
277998

278003 TGGCTGGACGGCGGCTCATGATGCTTAAAGCGGTGATGACATGATGAATTTGAGATTA |-----|
278040

278063 CTCACAGCATCTGGCTTACCGATATGTTAAACGCTTCTCCCTATCTGGCCGGAAGCTCAG |-----|
278092

278123 AGGCGATGCTTCTCTACTTAACCAAGAAAGTGAATGCTGACTGCGGAAACCAAGAAA |-----|
278183 CAGGATGATCCACCATTTGGCCGTGGTAAAAAAGCGCGCTGATGGAACCCGCTGTTCCT |-----|
278216

278243 GTGTGTGCTGTGGTTGAACCTCTCGGACG |-----|
278278 TCGCGTTTACCTGGTACCGGCATTCATTTGCCCTGATGCTGGAACCGCTTATGCTGATA |-----|
278328

278348 GTGCGATGATGTTTGTCTATCTGGGCTGCTGCTGATCTGTATTTTGT |-----|
278496 GGTACAGATCGCGGCAACCAACAGCCGTATGTTGTACTGGGTGACTGGCCTGGGC |-----|
278546

278556 ATTCTGCAAGCGCTCATGTTCTCCGGGCAAAACCTGAATCTGAGGAAAAAGGAGTTAATCC |-----|
278565


```
271822 CTGCCCCCGATCCAGCGCCCGGTGAACAAATCAAGATGCCATTGTGTAGCGCGCTGCT 1-----1 1-----1
271836 271836
271882 CCTGCGTAGCATGTGCCAATATATGTCTTCTCCGGACTGAATCATGTTCTTTTCT 1-----1
271942 CAACCGGCGTCATCTCAATATGACAAACCGGCTTAGCAGAAATTCACAGTACCGGCGCTGT 1-----1
271947 271947
272002 A ...
272070 TGCAAACTGACTCTGACCGGAGATGCCAACCAGCTGCACAGTTTCAATAAAACGATGATC 1-----1 1-----1
272130 --1 1-----1
272132 272132 GCTCAGGCTGTGCCCCAAACCAATTACGGGGAAACTCCTCGTTCAGCCCCGAAGTTTTTC 1-----1
272190 AAGCA ...
272388 CAGCCTTGCCCAAGCATTGATTCCTTTTGCGTCATCATTTTCAGTATCAGACGCTGCAT 1-----1
272448 CGATGTAGACTTTTCTGTGTAGGAGCAGCAGCTGGCCGAGCATATGG 272438
...
272722 GGCCTCAGGGCCAGCTCGTCACCATTTTACGACGGGTCTGAAGAGCGGTGCTCAGCAAG 1-----1 1-----1
272782 TCGCTCAGAACACAGCGGTTTATTCAATAGTAACAATCTCCACGACGGCGCACCTCAGC 272786
272842 1-----1
272843 CAGCGCTTTAAGAACTGGGTGTAATTAATCTTCTCGGGGTCTCAGTACACACAAGAAAT 1-----1
272902 GCGGTCAAACTGCTGTGGCGACGTTATAGACAAAAATTGGGGATGCCAGCCGTAATT 1-----1 1-----1
272920 272920
272923 272923
272962 ATCCGTAAGAAGAAATGGCGGACTCAATGGCATACCAACGGGACAGGCGGTGGT 1-----1
273022 CGATCCATAAATGCTGTGCTCGGCAAGCTTGA 273001
...
273309 TGACGGGCAAGGGGTGCGCTCCGACAGGGCATTTGCCACTCCAGTCCGTCAGCGTTACGG 1-----1 1-----1
273337 273337
273369 CGATTACCTGTTCAATATGTGAAGCTTGCAGGATTAACGACAGGCGAAGACAGGTTAA 1-----1
273372 273372
273429 TAAAGTATTACGGGTGGTTGCTGCTCATCAATCAAAACGACGCTCGCGTTGGTTAA 1-----1
273470 273470

273489 1-----1
273492 CACGACGTCTTCTTCTTCATCATCTGCGAGATAGATCAGATGCGTGGCGCTG
...
273782 CCGAGTACTTGTGTACAAACAGAAAGCGGCTTTCGGGTACGTCTTTCAGCGATGTCA 1-----1
273832 273832
273842 AACAGTCAATCGAGCGAAACTTCACCCCTGCTGACGGGTACTGATGATCGTACCGATGAG 1-----1
273902 AGGGTACGGCGCATTAATCATATCTTCGAGATATCAGACATTAATACGCTTAATTCAGA 1-----1
273909 273909
273962 AGG ...
273986 GCATTTACATCCGTAATGATCTCACCGCAACCGGAGAAAGCTGTTTACCGCTTGG 1-----1
274036 274036
274046 CGTACCATATCTTTCAGACATCAGACCGAGCTTAAAGCAGCAATTA ...
...
274133 CCCCCTGACGAGCAGATGTTGTTTCAAGCAGCACCAGGTTGCCATTACATCGTTTGGC 1-----1
274183 274183
274193 GTCTGAACATTACCAAGCCCGTCCGCTTCATTAACAGGGGCGCATCACAGCCAGTTCC 1-----1
274253 CAGCGCTTCTCTTCTTACAGATACGTGATAGCCCTTCTTACGGGCTCCACGGCT 1-----1
274255 274255
274313 GCGAGTATTTGCGCCTTATCCGCGAGGATATCGACGGAAATTTCCGGGCTCGCAGTTAC 1-----1
274373 GGCATCAAAACAGGGGTTTAAAGACTCAGCAGCTGATGCTGCAATATATAGTTGC 1-----1
274394 274394
274433 GGCATGACCAGACGATGTTCCGGGATGATTAAGACGGCCATAGCGGAAGCACTGTATCA 1-----1
274443 274443
274493 1-----1
274494 AAGCGCCAGAAATCCCATGGCATAGATACCCGTCACCGGCTTCACGCATACCGGCTG 1-----1
...
274637 TCCATAAAGTAGCAAAAGTAACCTTTTCTGTCAGCAGAGTTAACAGTCCGCGCGCTT 1-----1
274687 274687
274697 GCACCGGTAGTAAGGCTGACCGGCTCGTTCAATGCTGAACGGT ...
...
274774 ATGAATATGCTGATGCCGTAGTCTGAATGTTTCTGTGATAAATGAAAGAGGAGTT 1-----1
274824 274824
274834 AGGATCTCAGGTTCACTCAAAAGAAATATCGCAACGGAAGAGATT ...
...
274978 TAAACCTTCATTAACCAATTTTCTTGTTCATTTTTTTTAATTAATCCGGAACGCGATGC 1-----1
275028 275028
```

```
...
269431 ACCCTCCACATCAATCAGCTCCTGCTGCCACCGGAGAGATTTTGGCGCTCAGC |-----|
269491 TTTAGCTGATTCGATTCGAACCAATTTACCNAAAAACTAATGCTT
...
269562 ATACAAGTAGTCTGGCTTTTCTGAGCGACCGGTCGCGAAGCTGAAAAGGCGAGCTTC |-----|
269612 |-----|
269615
269622 TGTATGCAAGTATACCTGCTTCAGCTCCAGCTCCAGCTGATGCTGTTTGGATCCGGCG |-----|
269678 |-----|
269682 TCAGCTCCGTAATATTACGATTGGCGTTACCGCTGACATGAGTTGACGGCTTCTGCT |-----|
269683 |-----|
269742 CAAGCGCTGCTGAGAGCCCGGGTGGTTGCCGTCGGTTTTCACCTCGAAGCGGC |-----|
269744 |-----|
...
269891 CCTTCGCGCCCGACACTGAGCTCGGTGCGACCGCTTACTGAGTTGAGCGGAGACGAGA |-----|
269941 |-----|
269951 AATAATGATGACACTTCTGTAACCTGAACCCCAATTGATTTCTTAATGATGCTGGGT |-----|
270011 CACGGTAAGTAAATGATTCATCCATCAGCGCTTGGTCTGTTAACCCTAATAAAGAGAC |-----|
270035 |-----|
270071 CCTGTCAGAAATGCCAGCGCCCGAGCGGATGTGTGGCGGTTGACACATTTGCTCCCTG |-----|
270097 |-----|
270131 CGACTTAAACAGCGCCTGGGTAGAGTTCAAGATCCGCAAGCGCTCAGAAATGCTCACCTG |-----|
270170 |-----|
270191 AGTGACCATCAGCGCGCTTTTGAATGACCTCTATCT
...
270252 ACAAGCCCGTGTGCCATATCAAGAGCCAGCTCTTGCCCATTAACGGCATGATGATCAATA |-----|
270302 |-----|
270312 TGCTCACAATACAGCGCGGTCTAGGGCAAAATCAGCGGGCGCAAGACCGAAACGAGCAT |-----|
270372 TTGATGAGAGTACCCATCGGGCTGTCAATACAGCTCAGGTCCCGGGGAGCGGAGCGAG |-----|
270402 |-----|
270432 ACAAGCTTCATTAAGATCGTTTGGCCCGGATTCGAAGCGCAATATGCGGTCTGCGAGGGA |-----|
270434 |-----|
270446 |-----|
270492 TGTTCCTCCA
...
270505 GGGTAGCCAGCTCGCGCATCTGAACCAAGCTCAAGACATCTTCTGTTCCAGCGTTGGCA |-----|
|-----|
```

```
270565 TCATGCAAGAGTAGTCCCGACCATGATATCCACCATTCAGCTAA
...
270741 CTACTCATGCTGTATTACTAGAGCTGCTGACAGAGAGTGTCCAGATTTTGAAGGGGAGGGG |-----|
270791 |-----|
270801 ATGTCCTTCACTGACAGCAATCCTCCAGGCAAAATCACCAGAGAGC
...
270890 TTCCCGTGTGGCAGCATGTATAGCGTGGCTCCAGATTCCAGGAGGAAGAAAGAGATTTT |-----|
270940 |-----|
270950 CATCAGAGTACCTTCTTGTATTAATGTTTACAGCCCGCATACTGCGCGA
...
271007 AATTCTGTAAACAATATTCCTTTTACAGAGCAAGCCCTACAGCAAGGCAAGCATGTCGGG |-----|
271057 |-----|
271067 TCGTCATATGAACGCAACAATATCATATGTCGGGTACCCGACGTAATACAGCCCGGGTGTCT |-----|
271071 |-----|
271127 TCAGCAATACCGGGTTTAAATACGATTAACGCTGTCCACTTTAAATTCATCGGCGAGGAAC |-----|
271153 |-----|
271187 GCGATTAACCTCGGACCTGTCTGTGCCATTAAGACCGGGCTGATTCGTAATTCAGGTCAGG |-----|
271192 |-----|
271247 GGAGCGAGGAGATGACGGGTGTTAATCTTTACGGAATGAGCGACGATATCGGCAAGCATC |-----|
271285 |-----|
271307 CGGCTGCAATTCGAATTCATGAGATGCTCGCAAA
...
271350 CGTGTAAATCCTTCAGAGGACCATCTAGACCGGAGATCAGGCTGATACCGGCGCCCGCA |-----|
271400 |-----|
271410 TGATGCAAAAGGAATGAAGCCAGTCATCATCTAGGCTGCAAGCCA
...
271462 GCCACAGGAGATGCGCCAGAGACAACAGTCCGGCTGCTCAGGATAGCCGGCGGATCTTT |-----|
271512 |-----|
271522 CAGTCCGCTTACAGCTCTCGGGTAATGCGCTTTTACCTGTTCATCCGTTCAACAAGAAC |-----|
271582 AATACCGCTGTATACCATGCGGCTTTCAATGACGTCGAGAGGCTGCAACGCTCAATTCAGC |-----|
271612 |-----|
271642 ATCCCGATTAATGCTGATACGCTAATGCCAATGAGGTTTCCCATGTCACGAGGAGCCTG |-----|
271702 GTGACGATTAACGCCCGAGCGGAGCCAGCCCTGACAAGACTGGCCAGTACAAATGGGCT |-----|
271711 |-----|
271724 |-----|
271762 CATCACCGAAGCGTTCCGCGCAGAGCAATAGCAGGCTGTGTGACTTCTTTTGCAGACCTCT |-----|
271771 |-----|
```

```
266279 GGAGCAGCTGTGTACGGGCTGATCGGACCTGCCG
...
266381 AACCGATTGATGAACTCCTGCCATGGAAGTAGTACTCACCATAAATAAACCGTCAATA
266441 CGGTCTCCGTTGACGCTTACTCTTTCTCCGCTGACGTAAAGCCACAGCCAGCCGAGAGA
266501 GGATTTACCGCCTTTCCGGGTAATCAAGAGACCTCAGGGAGCTTTCATCTGCATGAAATAC
266561 ATCCCGTGTCAAGAGCTCGTGGCGTAACAAATTTTCCAGCGGGGCCAGTGCAGCCGACG
266621 AACACCGACCAATATCCGCCATTTGTAAGACGGGAGCTCCACCCCATTTCCGGAGAAAT
266681 GTGCTGTGGGATACAGGGGCGAGATGGTCCGTATATTTGCTGATGACACCTGTGGAT
266741 GACC
...
266879 CTCGCTGACTTCATCGCGAATGAAGTGCAGGGGGCATCAGTCAGGACAACGTTCCCTC
266939 AGCTGGCGGATACCTTTTTCAGCAGGGGAAGGTGAAGGCAAG
...
267050 ATGCGCCGAGATTTCTGCGAATATCCGCGTCAACATCTCTTCGACAGGAGGAAGCTGCAT
267110 CCCGGCCAGCATTCGCAATTTTTTACGGAACGCTGTGGGTGCGAGCTTCATCTGCTC
267170 TTCCAGATGCGGATACGGTTCTGA
...
267464 ATGCCATGTGGCATTACCTTTCCGGGGCCAGCGGAAGCTTCCCGGTGACAGTCGCGGGGT
267524 ACACAGCCATACGCCATGTTGTCTCCACCGGAGAACTTTTAATGCGGAGCGGACTTTTAT
267584 GGTGAAGACAAAGACGACCGGTGATGCCATGCAAGCGTTAAGTTGCTTGGAGACTAG
267644 TGTCACTGTGTGATACCCCGGCGCATATCGAGGGGTTTACGACAGGAATACGTTATC
267704 AGCGTTCAGCATGGCGAAGAGCCCGAANAACATCGTTAAGTGAAGAAGGCTGGCAGC
267959 GGGTAAGCGCCGTGGCGGCCATGGCGGCAAGTATGTTTTTTCTGATCAAGCGTCAATC
...
268019 GGGTTTTCAGATTGATTATCTCCGGTAAGTTGGTAATGATGTCCGTAATCTTAATCG
268079 ACTGAAGATCCCTGTGAAGATGACTTCCCGGTAAAGACACAGATGCCAATCAATGAGC
268139 GGAAGCATGCCAATCGTGAATGATTGT
...
268238 ACAGAGTGAATATGACAAACCGGAGTTTTCACCCCGGTCTGAGATTTTACGTCAGGA
268298 CTTTTAGCGCTGTATATCTTTATATCAGTTTCATCAAGGAGTGCAGAGCAGCTTCTTAC
268358 CG
...
268375 AGTCACTAGCCGACGACGAGCGTCAATTTTCTTGAATCCGATGCGTACGGAATG
268435 ACACGGCTGTTTTTTTATCTTATGAGTAGAGACACCTCCCTGTGAGTACACCGTTTCA
268495 GTTACGCAATTCGTGCGAGCGCTTTCATTCGCGGAGGCTTTTTTATCCAACGTGC
268555 CCGGATTAACGAG
...
268793 TCAATAGCCTGAATAATGAGGTACTGCTCATCTTTGCAATTAACCGTTGATACGCTTAC
268853 GCTTCAACCGGATCAATGTGTTTTATGCTTGCATAAATTACTGACGACATCTGCGTCA
268913 GTGCATTTGCAAGACAGATATATCTCATCTCTCCAGATA
...
268970 CGTTGCGTAAAGCCTGACACAATCAGTTTTCGAAAGGTAGCGTATGTAATAACGATACCG
269030 ACCAGCGGAAACAGACCATGACGTGTGGCATATGTTTTCCG
...
269078 ACGGCTTTTGGCTGATGGCTTTTCCGTCAGCTCCGTTCCCGGCAATTAACGCGTA
269138 ATCTTATCAGGGAATCTTCAATCTCAGCTAAATCAAGTCCAGCAGAGAGTGTGATAGC
269198 TCCGCTTTACTTCAATGCAACAGCGTTTCCAGCGGCATTCAGTTTTCGCCAGACA
269258 CCCGCAAGTATGATTCGAAAACGCGGCTGTGTCCGGAATGCTGCGGGCATGTGCC
269318 TGAACCGTGGCAGCCGTTTTT
```

```
263238 GGCAGCAACTCTTCCAGTCGCTCTTCCGGCCACTCCGGTAGACGGCTCAGAAAGCTCTGTC
263298 AGCCAGGCATGGGCTCCAGACCATTACGTTTGGCGTTTCCAGAA
...
263467 CGTCACTCAGNAGCAGCTCAGTTCATCGCGCGGCTCAGATATATTAATCGCTTGT
263527 TTAGTGTCCGTCGGGCGGACACAGCGCTCCTGCTTTCAGGCCAGGCAAAACAGGTCAT
263587 TCACATCGGCGCGGAATACCGTTGTCGCCAATTGCCGTAATTTTTCACAGAGACGTTAC
263647 GGATGAACCTTCTCAATGGGATTAAGACCCGGGATTTTCTCAGCGCCATGGCTGCACGGG
263707 GCTCCCTGCTGGCCTTATACAGATTGGGAAACCGCG
...
263853 AGCCAGGCTCAGATATCTGAGGGCCGTCGGCTCAGATCAAAACAGACGACGCGC
263913 GGACCGCTTCTTCTCGTAAGGTAACTGACCGCGCTGATGAGCC
...
264149 CCTCTGTTCAATGGCTCAGGCTTCGACAGACGAAAGCGTATATCGCGCGCTTCC
264209 GGTACACACCTCTTCTGGCGCTGGCGTTGAATATATTACCTGTGAGATTAAGCGCTGATGA
264269 GCAGAAAGAGCCTATGGCGGCTTACTCCGCTTTTATCACTC
...
264321 AGCAGTCAGTCGCCAGTTCCGTAAAGTTGAGTCCGCGACAGGTAAACATGACGCTGGAA
264381 AATCCTTACCAGAGCTGCTCAGCGTGTGATCCGTGAATGACCGGAGGAGGACGATGA
264441 TCTCGCTCCCATCAGGCACTGGAATCTGGCTGGTTGCCGGGTACTGATATGCGTAAAT
264501 CCTTCAAAGCTCTGGAGAACAGGTACAAATGT
...
264592 CACGATTAAAAATTCTGTGGGCTGATGCTGATGTCTGTGCTGTTTACACAAACGCTGGA
264652 GGAAGGCCAGTTTATCTGGCCTGGCGTCCGTGACGGTAAGGTATCCATTACCCGCTGCA
264712 GCTGGCAATGCTCTCGATTAACCTGAGACTGCGTCAGCCCAAAACATCCGCTTAATGC
264771
264772 ACTGACAAATGTTGTAATAAATGTCAT
...
265059 AGCGTAGAGTTGMAACAGGCTGAGGCCCACTGMAAACAGCGAACAGAAAGCGATCGTT
265119 ACAGTGGCGGTGAGAGATGACCCGAGGTTCCCGCCAGTTGCGACAGTCACGTCATCGTC
265179 GTCCGTTACCGGACACATCTTCCCGTGAAATACACCGTCTGGAGCCTGAAGAA
265257 AGCTGATTTATCTGGGGAGTCAAGTCTGAGACAGCTGGAACGTGATGATGAGCGCCTGA
265317 AAGTGATCCGACTGTTGCGGTAAAAAAGCCTGTACAAAAATGTGACTGTATAGTTGAG
265377 CGCCAGCGCCCTCCCGCCGATPAGAGCGGCGATCGGGCTCCGGATTACTTGCCCGCG
265437 TGTTAAGCGGAAATACTGCGACACCTGCTCCACTGTATCTGTCAGAGTGAATCTTTGCC
265497 GCCAGGGCATCGAACTGAGCGGGCATTTACTC
265569 TGAATGATACCTGTACCGTTACGTGATGAACACCCGCAAGTTTACACTGACGACAC
265629 CAGTAAAGTGTCTGGCACCGGGCAGAAAAAAGGCAAAAACAGAGACG
...
265963 CAGTGAACCTGTACGCCATCGAGATGAATATACGGGATTTACCGAGTCAAGCGCTTGC
266023 CGTCAGCGACGAGGAAAGCAAGTGTACTGACGTCACTGCATGATGATGATGTGAGAA
266083 GAATGTACGCTGTGAAAAAATCA
...
266159 TATTACAGTGAATGACGGTCTGCGGAGGCGGATATATATGCTGGGAAAGAGCGCTTCGT
266219 GCAGTCTGTCTCGGAAAGAAACCTTATATGTTCTTTGGCAGCGATCACAGCGCGACGCT
266258
```

```
...
259867 CTGGAAGATTACCTGAAGCCGGATGTGAAAGCCCGAGCTCAACTTAACGTTAAAC
259927 CGAGGCCCTGACCATCTACCTTCACCAAGTGAAGGTTAGCGCTCTCTGTAAGAAAGAGC
259987 AAGTCTATGCCCGCAAAAACGATCATTTGTGGCATGTTATGCTTGACCATGTTCCTGAC
260047 CGTCTGGGTGTTACAGCATCACCGTGTGAATTCAGGGTAAAGCTTCATGTGTGTCGAAT
260107 TGGCGCTTCTTGACAGTGAACCGTAAGAAACCGCAGCGGGGACACAGTCCCGCTA
260110
...
260614 GGTTCACATTCAGATTCTGTGATGTTTGTAAAGTGAAGATAAAGCATTAAT
260674 GCAGAGGTATTTGCTAAACTATACATATCATTTATATTAACAAAGAT
...
261103 TTGTGTTATAGTATCATCATGAAGAGTGCAAAACACATCATCTTTTATTGTGGCATTAAT
261163 GGGGGTGGGTGAGGTTAAATATTAAGACGATGCTTAGTGTCGTAATGAATTTT
261223 TCGCCATGACACAGATCGGTGTTT
...
261302 TGGGATATTTGTGTGTTTTCTATGAATAATAATAAATCAAGCAGACAGCTGTT
261362 TTAGTTTTTTTGCATCTTCTTATATGTCTGACATTATATATCAT
...
261515 TGTAAATATACTGGAATGTATATCCGGCTGATAAAAATGCCAATATATATACGCTAATG
261575 CTGTTTGTAGGTTTTTATAGATGCGTTATTATTGATTTAA
...
261836 TTGGGTTGACGAGATTCCCTTTTGTAACTGTATTGCAVTAACGTGCGCTTTTG
261896 CGACCAACCAATGTTCCAGCAGCTGAGAGGGAGCGCTTCGGGTGACTGCCACCAATGT
261956 TTCAGCATCTGAGAAAGCGCTTCGGGTGATTGCAATATCTTCAGCCTGAAGAG
262016 ATGCTTTATGCT
261973
```

```
...
262107 GTGCTCCGACAAACCGACTACTGTGTATTCGCCGACAGACTGGCATAAAGCGCTTTA
262167 AATCTCGCTCTTAAGATATTTTCTGACTTCCGATCCGCTCTTATGACTGTGCGAC
262227 GGCATAGACCATTTGAATTTGTGCTGTGGTGTGCATGTGAGTTTGAAGCTGTGCG
262287 GGGCTCACTTTCGTCAGCGTTAAAGCTGACAAAGTGACGCTGCGACTACTCAACGAA
262347 CCGCATTAACGTGCTCTTTACGAGATGTTTTTTAACAATTTGAAGAAAGTGACACAGCTG
262407 AACGTTTTTTTCAACCTCATTTCTGTATGCTACTGCGCGCTCTCCGACTGACTTAA
262467 AGCTTCGTACCAACAGAAATCGCATTTGTTTTTAAGAGTTTTTATGTAGTATACGAA
262527 TCGCGATATGTCTGACAGCAGATATATCGGGGGCGACGTTTAAAGGTACAGTTAATTC
262587 TGGGTGCGTTAGCCAAACGACTTGACA
...
262643 ACCGTGACGACAGGTACTCACTACCGAAGAAAGTAACATAAACCCTGCGACATA
262703 ATATTCCGGCACTCATATTATGACGCTTTTTCGGCGCATCATCAGCGATACGAA
262763 TGATCGCAGCTCGAGATGTTGATTTATTTCTGTTGCCCGCGCGAGCCGACAGCTGCT
262823 ATCGAGGCGGGTTCTGAAGAACTGAGGTATTTGTTCCG
...
262875 ACGCATTTTTTCAGCGAGCCACTGCGACGGCATACGATTGATCAACCGTCTGACGATGTT
262935 GATTTATTTTGTGACACAGGGGCGACCGCCAGGCTGTATCGAGCGGTTTCTGAAGAA
262995 CTGAGTCGGGTAAATG
...
263068 GGAATGGCTCCGGGTGATTCAGGTTCTGACTACTGTACGCCCGCGGTGCTGCTC
263128 TGGCTCATTTCCGTGTCAGGCGGCTTCACACTCCGACTGCGGGTTATAGCCCGCTGCACG
263188 AACACCTGATGCGCAGG
...
263141 263148
```

```
256730 -----|
      TCGCACGCTTCGGCTGCAAAATATATGCCCCGAGAGAGCTGGCTACCATTCAGATTCAC
      256734
...
256807 AGTTTGGCGTGGATGTTGATTTTGGACATCTTCCGTTTGAGCCGGACCGGCGCTTACG
      |-----|
      256857
256867 CCAGCAAGTCGCACGCCCATATATCAGATCAGAGACACAGCAGCCACTGAATGCCGACTCCGG
      |-----|
      256927
      AAAAAAGCTCCGGCTGATGACGCTTGCAGTAACAGAGCTCCCTGTTGGAGGCTACAGACT
      256930
256987 GGTCCAGGAGACTGGAGTGGGCAGTGGAAAGCAGCGCTGGGTAGAAAGATAGCGCCGCTTGG
      |-----|
      257034
257047 AACACTGGCAGCAGATGTGCAGATGGAACAGAGTTTTTTTGGCGTCGACCTGCCACTGCTTG
      |-----|
      257107
      CCGGCGTATPACGGGCGCTGTGAAGCGGGGAGTCCCTGACAGTGCAGGCGCTGGAACCGCTT
      257133
      257152
257167 ATCTGCTGGCCTGCAGGGAACCCGCTGAGCTACGCGATGAA
      |-----|
      257350
      ACCTGGCCCTTGAAGCCTTGGCTACAGCTGATTTGAAGCGCAGTGATGGCGGCGCTACAGGC
      |-----|
      257410
      TGTGATCCCTACGGCCAGCTGCTGCTGCCAGAGATTAAATTATGGGCT
      |-----|
      257472
      GCGAATGAGACAGTGCCTTGCAAACCCCGGATCAGCGAATTGGTTACGACAGCCACTT
      |-----|
      257532
      GCATCTATCGCTCCGGCGGACATGAACCCAGTATCCCGATTAATTCGTTCTTAGSAG
      |-----|
      257592
      AAAATAAATGAAACATATATCAAGCAACCGCTGCGTGCAGTGCAGGCGGTCGGTCCGCT
      |-----|
      257652
      CAGGTAATAACCGGCTTCTGGAAGCCCTCTGCAAAATCCATGCTGACACCTGGCAATTAG
      |-----|
      257712
      CCGTAGTCACCAAGCAGATTCACACAGAGAAATCAACGATTCCTCACTGAGCGCGAG
      |-----|
      257772
      CACTGGA
      |-----|
      257868
      CACTGAGTGAATAATTTGTAATCTCGAGCTGATCTTTTGGAAGGCGGCGGCAAC
      |-----|
      257928
      TGAGTGGACCTTCAGCCCCGAACTGGCGGACCTTACCATATATGT
      |-----|
      258027
      CCAATCCGACTTCCTCGTATCAATATAAACTGACCTTGACCATPACGTGGGCGCTCC
      |-----|
      258077

258087 TCAGAGTAATGAGCGTGTACTTACGTATATGCGGAGAGACGCCCAATGGGGGTTCACTA
      |-----|
      258147
      ACCTGAAAAGTGTGGAAGTCTGCAAAAACATCATCGCTTTTATGAAAGACAGGCGCATGC
      |-----|
      258207
      TGGGTAAAGTAATCAACAATCCTCTGTA
      |-----|
      258236
      CAGGCGCGGACAGCGGCCCTTGTTTTATCAGAAGAAATPACGAAGTTGGCGCTCACC
      |-----|
      258296
      GCGTTATCCAGTTGTACTTACAGATGCCGATTTTCAAACTCCA
      |-----|
      258544
      AATCTACAGCTGCTTTTTTAAATATATCTGCTCAGAGCGAGCTTCATTTAAGCCCTTAC
      |-----|
      258604
      GCAGTTCAGAAATTTACAGATTCAGATTACAGCCACCGTGGCGGAAC
      |-----|
      258732
      CTGATCGAAAGTTGATTTTCAAGAACCGTCTGACAGCTTGGCTTGAACCTTTAGAG
      |-----|
      258792
      TAAGCTGGGTTTTTTCGTCATATTAGCTCCTCTGATGCCATTTCTATTTCAGA
      |-----|
      258872
      TACCTCAGGTTTTCACATCATCTGGGTTTTTACTTCAGATTTGCTTTTGTGATCGCTGCG
      |-----|
      258932
      CCATGGCTCCACATCTCGAAAGACCGCAGAGCTGACCAAGTC
      |-----|
      259084
      AGTATGTTAAAGATPACAGGTAAAAAACAGCGGATCAGATPACGTGGGTTAAAGCTTATTT
      |-----|
      259144
      TCATPAGAAGTTTTTAAAGTCAGAAATAAATAATGGTGTGTATATA
      |-----|
      259459
      ACAGTAGTATTTCACGATPACAGATATTGATGAGTACTTCAGAGATCGGTTGAGAGATAAA
      |-----|
      259519
      ACTGACCGGGTAATTTGAACGTGATGTTAATCTTACCATATGTTACATTTGACGTTTCT
      |-----|
      259579
      TCTAAGTCGCATCCGTTTTTATACAGTAAACTGAGAACTGCTGCTTCAGAAAGTAAATGTT
      |-----|
      259639
      GCAAGCTTTACCGACGTTTTTGGCGTTTTTGTGATGCGAAAAAAGGTGGCTGATTTTGC
      |-----|
      259699
      CGCCAGGGTGCCTTTCAC
```



```
249737 CGGGAGGGGGGAGCTCATCTCAGTTAGCGGCAAAATCGCATTCACAGCCATTCGGCTCTC |-----|
249797 TCGATTCGCCCATCGGGCAACAATGATGTGGCCACAAAATCCTGGC ...
250072 TGGCTGAGTCATGACATCAGTGGCTTATCCGGAACAATGATCATGTGTTGGCCAGAG |-----|
250132 AAAGCATTGCCATGAACAGTTTGTATTATCCCGGAAAGTCATTT 250122
250397 CAGATTATTAAAGCAGCGGTTATTCATGGAAGGCTTATGTGGCGCATGGCCTATGA |-----|
250457 AGCAGCCTTTGCTCAGAGTGCATTTCTGTCTGTCAAGCACTGATA 250447
...
250679 GATTACGATTTTGTGCTATCATTTATATGGGAGTTTATTACTTTTCATACGCCCAAC |-----|
250739 ATCTAATTATTCAAGATTACCTGGTTAATATTACTTAAGATTA 250729
...
250843 TATCATCTTCTTTATTATAAAGTGAAGTATTACATTCACATACCCCAAGCATGATAA |-----|
250903 ATATATTATCATGACTTACTGTCTGTATTAGAGCATGTTCTGACAA 250893
...
251933 GAAAAAGCAATCGCAAAATATATCAAGAGATGCAAAACATGTATTTAAAGCATATGT |-----|
251993 AATACCTTTATTCATGTCTCAGATAAGGATATTACTAACAAAC 251893
...
252526 TTATACCAAGTCAAACTTCACATCATCAAAACTTTATGTCAATAAAGCGTTATAA |-----|
252586 GTTACTGATCTTTTATTATTACAAACCACTACATCTTTTATAA 252576
...
252686 GGAATATATTATTATTCATTTCATCTCTGGAGGCCCCCTCGATATGAAGCATATCA |-----|
252746 CACAGAGCCCTGCGGTGTACAGATAAAGTTTGCAGACCAAGTCAGTCAAGAGCG |-----|
252802 252802
252805
252806 -|-|
252807 -|-|
252810 -|-|
ACGTCCGCAACAATGTATGACTGTATACAGATGCTCTCCCGTATGACTGAAGACAG
252807
252810
```

```
252866 GCATGCATGAATAATACGTGCTTCCCAAGAAAAAGGGAAATTAATCATCTGCATGACCC |-----|
252883
252926 GACCAGTGTCTAG
...
253073 TCTTTCTGCTGATTTCTGTGATAATCCCGCAATATGCAATTGCTTTCTGCGCATGAT |-----|
253123
253133 AGCCCCGGAGACCGGAACGCTCATGACGAAGAAACAACATTCGATTATTACAGAGAAA |-----|
253148
253193 CTCACGTGMAAGGCATCATCTTTATATTAACAGAGAAATCCGCTCTTTACTGTGTTTT |-----|
253245
253253 ATCTTTATTATTACAGGGCGTAAGATTGCTCAGTACCGCGTTTACCACAAAATAAAAAAG
253313 CTGGCAGGCCATGCTGATCTCCGTTTTCAGCGTCTGCACGGGAAACACAGCTCACAC |-----|
253342
253373 CCGTCATCATGTGGGTCTGCTCAGTGTTCAGCGTCTTTTATCCCGAAGAGAAACCTG |-----|
253402
253433 CCACCTCTGTCGTCTCATCCACCCGGTGTATGTCGGCGGAGTACGAATCATATCAA |-----|
253464
253493 CGCAATTAATTGACTCTGATTGCCATAC
...
253542 AGTAATTTTACCGCAGCAGCGGGCGCAGCGCACCTTCAGCAAAACCTGACCTGTGC |-----|
253592
253602 AATAAGCAACGCTCGAGAGTGTTCACACAGGATTCGATTTTTTCCCGGCGTCAGGCC |-----|
253610 253650
253662 ACGTTGCATACCGGTTCATCTGTGTCCTCCAGTACACTGTGGCTGGGATCTTCTG |-----|
253685
253722 TGCCTTGAGGTCCGGTTCATCGGAGAAACCTTCAGCCACGACATCGTAACCGTCTG |-----|
253729
253782 GAGATATGGTGCAGCGTTCTCCGCTGTGTGATTGAGCGTCTGCACAGTGGCCATGAGAA |-----|
253816
253842 CTTACCTGCGTTGCCCGGAAACCATGGGTGGGAAGATGCTTCTACCTGGGAATGAA |-----|
253874
253902 ACGCACTGGAAGCATACCGGAAAAAC
...
254056 AACTGACTGAACAATACCGGTACTTCCCGGATCTGGCTGACTTAAAGAGACCTATGG |-----|
254106
254116 AACTGACTCCAGAGAAAAAGACAAACTATTACTGTTTACCGCTGCACGTGTCAGAGC |-----|
254173
```


246778 -| |-----|
TTACCTCACAGCTGATACAGAGGGGTTCCGGGTCTCTGCGAGCCGATTCTGGCTGTTCGG
246779 246819
246838 CAGCTGCATCGAGAGTCACAGAGTAGCTGGGGCTCTCTGAGAAGGTAATAGCAGCT
246898 TTCCCTCGGGCCATTGTGTCCGCAATTTGACACAGATTGGCGTTAATGCCATTTTCAA
246916 246938
246958 GAGCAAGTTTGTAGATGATATCCCGGTTCCAGAG
...
247141 ACCAGAGGTTGAGGCTGTACGCTTACGATTACGCTGAACTTACGCTGAGATCGTCGAAG
|-----|
247201 GACTGGATTACAGCGCAGTAGCTATGCAAGAAATGCTCATTTTGGC 247191
...
247303 GCTTACGCTCATGAGGTTTATGCCGCCAGTATCCCTTAACATAGTGAGATCGTTCGGG
|-----|
247363 GTTTGAACCCAGATATTGCTATTTCAGTAAGCCTGTCCGTT 247353
...
247495 TTACTGTTCGTTGACCAATTGAATCTACACAGCGTATTTGACAGAGATCGCTGAAA
|-----| 247545
247555 AATPACCGCTTAATCGAACAATGGCGCAGACAAATTTCCCATTTAAGGATGTGCAGTAT
|-----| 247576 247598
247615 TTGTGTATCCAGAGCGGTTATTACAACTGGGTACAGC
...
247729 AAACATATGGAACCCGGCGGCTCCAGAGGAGCTGGCAGAGAAATGCGCATCTGTGTC
|-----| 247779
247789 GTGACCGACTGGCAGCTTTCGTAAGGAGCTAAGGCTACGCTTAAGCAGAAACGCAAT
|-----|
247849 TCAGGCGGACTGAGAACTCGAACCAATCTGCCAGTTGGCCAAATCTGCTGAACCA
247853 247907
247909 CCTTGGCTCTACAGACCAAAATCAAGTCTGGGTGGCGAGCTGACGATGTTG
247912
...
248031 GCCATGGGAGAGCGCATGACAAAAGAGCTGACAGGTAAAGCCCTGTTTATGGCGCTCAGG
|-----| 248081
248091 AGCCAGCGCCACCTGCGGGGCTAATCCAGCACCTCTGATCGAGTTCAAGTACTGCGCA
248095
...
248173 AGCAGTTTGTCTGAAAAATCAATGTGCGTAAGGTAATGTTACGACAACGCTCGCA
|-----| 248223

248233 TGGAAAGCTTCTGGGGAACGCTGA AAAAATGAGAGCCTGAGCCACTATCTTTAAGAGCC
248249 248278
248293 GGGATGAGGCCATCTCAGTAATACGGGAATACATTGAGATTTTCTAACAATCGTCAGCGTC
|-----| 248341 248350
248353 GTCACTCTGCTGGGGAAATATCTCCCGGCAACCTTCAAGGAAAAATATCAT
248355
...
248579 GTTATCAGATCATGACTTCCCGGATAAAGCTTACTACTGATCCAGGCTAACGTTCTCA
|-----| 248629
248639 AAGCAGACAAAAATGATCTGCGTGAATCTCAGGGATGGCTGAGTCG
...
248825 TAAATTACACAATGATTACAGATGTTCAAGATTGCCGAGGAAACAGAACCCAGCTCAGAA
|-----| 248875
248885 CGGATAATGTTCAGGCTTGGCTCAAGGGAACATCAAGGCTCCAAAT
...
249005 TGAAGTTGAAAGAGCTAAACCTCATCTGGAAGACAGCTGGCAAAATCAATCGCCATA
|-----| 249055
249065 TACTGTGATTAATAATATCCGGTGGCCATTCACCTGGCCCTCCGCCCTTTATGTCACAC
|-----| 249122
249125 GTTCAGACAAATTCAGATATATTATTTCCCTGTCACACCATATATCCCATGCCGAGACAAAT
249127
249185 CTAAATGACGCTACAGTAATAATCTTCAGGCTTCGGTTTGGCTGAGCGCGGTAGCTGTAAT
|-----| 249211 249227
249245 TGTGTCTGTTCTCTTATACAAAGACAGTAGTACAAT
...
249311 GACGTAACCAATTAAGATCTCTTCCACTATTAATAATGTGAATCTGTACAAACGATTCGA
|-----| 249361
249371 TAACAGAACTCTCTGCTGCTACAGTTAAACACAGCTGTGGCATGAGTAATACATCTTT
|-----|
249431 TAAATTAACCTCGGCGGAGAGCGCCATTAATTTTGGAGCTTCTTGAACAATAATTAATTC
249449 249465
249491 TGAAGATCTTCTCCAGTTATTATTCAGCTCAAGATTCAAGTCATTAATTAAGGTCAAATTA
|-----| 249539
249551 CTCTGATTAAGGTATCAGTTATGTTACAGATTAATGATTAATCCGACGACAAACAATATTT
|-----| 249595
249611 AAAATCTGCAATATTACCCGACATCAGAGATTCAAAACCG
...
...

```
244128 AATGACTACGGCAGTGAGAACGTCATCAAAAAATACAAATCTAGGGCCCTGCACCCCTCCA      245662
      244147
244188 GTAATTGTTTCATCCAGTTGGTAGTTGGCCATATAGCCTTCTGTGAAGGCCCGTTTGTTCG
      244220
244248 CGGATATGAGAGGCTATTGTTGATACGTT
...
244445 ACATATCTCAATAGCAGCCTTTTATATAGTGTATATCAGGCTACGAAAGTACAGCCTTTAT
      244495
244505 AAATTGCTCAGGTGATATTCGACGCTTTTGTGACGTGGTTTGTTCGCCCTTAACATAC
      244528
244565 CTGATATTTTGTCTGTGAGA
...
244656 CTTTTCCGCAATTCGTTGGCTCGGTTATTAGCATCAACAGCTTGTATATAGCCTTCGCC
      244706
244716 CCGATTACGGCTAAGCTTACAGCGAGGCCGATTAGCGGGATGT
...
244782 ATGCCCAAGAGAGCAGTTCCGCTACCCGGTTGACCGGCCAGTCTGTAATGACGCGCAAGC
      244832
244842 ACATGGCGAAGGTACGTTTCTGGATCAGTCATTCAGTTTGCACGTCCGATCAGGCTG
      244908
244902 TACAGTACGCGTCCCCCGCTCAACCATGATCAGAGCCGAAAGAACAGAGTTTATAGA
      244962
244962 CCCAGACTGACCCGCCGAGGAGCATTTTCAGCGATGTGTGTGATTTCCACCCAGCCA
      244991
245022 TCGTCCGATAGTACGTACGTGCGGCCACGTGTTAAGTGGCTACGCGAAGCCTTGGCC
      245026
245082 AACTCTGAGTGTGCGGACAGGGTCTTCATCTTTTCACGCAACAGCTTTCAGAGGATTC
245142 AACACAGGTTTCGTTTTCGCTGAGTTCAGCAAGCCGCTGCTGCGGCAATTCCTCT
      245164
245202 ATATCCGCCCTATGCGC
...
245413 GATGATGACCTTTCTGTCCGGGCTGTAAAGCAACACATGACAGTGCACAGCTTAC
      245463
245473 CTGCATTGGGTCATCAGAACATAGCCCAACACCGCCGCTCTT
...
245551 CCGGGGATATCAGCGCATGAGATTGGCGTCAATCATATAGCATGAGAGCGCTCTT
      245601
245611 CCAGCGAGACAGACCGCGCATGATCATCCACCGCCGACAGCAGCATGAGAGCCTCA
      245662
245671 GCTCCACACCTTGGCCGCCGATATTTCTGACTGGCGATACAGCGGGTGTCTGCAT
      245704
245731 ACTTGAGGTACAGCAGCGGGGACAGCAGCCCCGGTCCGGCGATACCCCGCTGATGGCC
      245768
245791 GCGAAGGTGACAGTGCCTGCACAGATGCATGCACTGATACAGGCAATGTTTTCCCGTA
245851 CCGTCCGATTAACCGGAAGGCGTAACCATATCACTCCAGCTGTTCCGGGATTCCTGCG
      245870
245911 CCAGATACCTCAGTGAACCGCCGAGTTCCGGGACACAGCGGCGCAACAGTCCCT
      245949
245971 TTTGCTACGGGGTAGTATTCAGGGAACGGCTTACGGGTGCGGCTTGACGCAACGGAC
      246028
246031 GCTGTACTGGCGGGTCAATACACCCATACAGTCAAGGTATTCGCTCTTTCTGAAGCCGT
      246033
246091 TCAGATCGGCTTCATTTGTGCGATACGGCGGAGACTTTTTCGAAACGACTGCCAGAT
      246136
246151 TCATCCGGCGGAGTTTATCCAGCTGCGCTGCACATGTCTATTTCCGGCTCCCGGTTGC
246211 TCAGCTTTTCCGACAGGGCGTGATCAGCGCTTCTGTTCCGCCAGCGCTGTTACAGA
      246237
246271 GGAAGATGTCGTCAGAAAGATGTGCTTCATTAAGCTGATTTTACCGGGCTTATTCGT
      246278
246331 GACAACCAAGATTAAGATTTTACGATGTGTACAGGAGGTACAGAGCCGCTTAAGCTGT
      246389
246391 CGCAGTGCATACCTTCCAGCAGCATCGCCAGCTGCGCTGCGTAAGGAACACTTTGGCA
      246415
246451 TCACGGGCTGACGGCCAGGCGCAAGCGCCACAGCTCCAGCCGTTTGGTCAGAGGCACAGT
      246473
246511 CCGTCACCGGTGACAC
...
246538 ACCTGACTCGCGCTGCGGCCCGGAAAAATGAATAACATAGCCGACATGGGATCGCTTTC
      246588
246598 AGCGCCGTTTGTACTTTCGACAGCCGCGTTGAAGCATTTTCTCATATCGGTATACCG
      246598
246658 GCAACCAAGCAATTTTGTCCCGAAGGTAAAGGATCATCGCTTCAGTTCTGTATCA
      246697
246718 GCAGAGTCAAGAGCTTTTCGCTGACATTCGCAATTTGAAGCGGAGCTCCGTCGCCGAGC
      246759
      246774
```

```
...
240024 CTGATTTACTGTATGAACAAAACAGAGTTGTACCGGTTATTGAGGGAACAAAGCTTGTCT
240074
240084 GTGCGACCTAACTTGCACCTGTGACGACGTTATGCTTACCATGTGACATATATGTTCTCTGC
240107 240116
240144 TACTTGGGAGAGACTAATGCTTTACCATGTGACGCTTGTATTTAGAAACTCTGTAC
240178
240204 AGGTGAGTCAATTAATTAATGCTGTATAC
...
240518 TACCATCATTTTATGTGTAAGTTAATATCTGTAGAGCCTTAAGCTGGCTCGCTGTT
240568
240578 TTACTTAATGCTACAGAAATTAATATAGCGGTGTGCTGTGCTA
...
240808 TATCAGGTTGCTGTCTGCTGTGCTTACCTCGTTATTAATGCGCGGTACGGCTTCTG
240858
240868 TACTCTGATTAATTAATTAATTCACCTTCTGTACTTTTATATA
...
240957 GGACTTCATGATCTGAAAGACGTGAATGAACACATGAGAGTCTGTTTCAACGTTTC
241007
241017 TAAACTGTTACCTTTCAGAAATTAATTTTGCTTACAGAAAGACACG
...
241081 TGTAGGCACAGTATGTGATTGAAATGCACCGCTTACTTAAGTAAACCCAAACGATGCTG
241131
241141 GCGGGAATTAATGCTGATGAATATACGTCAAAACAGGACTTCATATTCATTTTATC
241201 GGTACTACGTGATGGCAAGTCACAAACAAATCTTATCGATATCCAGGCAACTGGAGAT
241217
241261 ATCTGAGGCGG
...
241322 ACCCGTGCCTATGATACAGTAATTCATTAATGATTAATTTGCGGTAGACGACTTAC
241372
241382 GTTACGCCCTTAGCTATCTGCGCAACAGATCAAGAAAGAACACGG
...
241522 TTTCTGTAAGTTGTTGGAGSTGCTTGAGTTAATAGCAGAGAGGCGGCGCCTTTGT
241572
241582 CATCAGTGTAGGAGACCGAATTTCTGCCGTGTGCTCAATGAGT
...
241687 GGCAGAAAATGCCGCTGTGCTCATTTACTTGAAGCTCATGGGATTTGAATGGCAGCTAA
241737
241747 ACCGCAAGACTCCTGTGCAGCGGCTTCTGTATTATTCACCGCAGAG
...
242297 TTATTTCTTAATCAGATTAATATGCCAAAGGGGATTTGTAATCTCATACCTTACT
242347
242357 CTGCAAAAAGACCGCGTAATTTGGGGGAAAGCGTTTGTGATATGAATCTCCAGCCT
242387
242417 TATCCTGATCAATGGGCTTTTCTGTATCGGTGACCCCGATGAATGTCAAGATTTGAA
242477
242479 CCGACGATATTAACGGGCTACAGGAGTATCCATCCCTGTGATGTAATTTATCAACG
242479
...
242821 CGTATTAATCTTTCCTTAAGGACACAAATATGCTCAGAAATTAATCATTAACGATTTGC
242871
242881 CGGACGGAATTAATGCGGTTTTTATGGAACCTCAGAGCGAG
...
243037 GAATACACTGATTAATGCTACACCGGACTGAATTTGCTGAAGCAGTGGCAGAGCGCTGCC
243087
243097 GGTGTTCTGCAGCGCGGTGACAGTATTTGATTAATCGGGGAGGT
...
243479 AGAATCCAGATATTTTCAGAGAAATGCTCTGATCATGACAGAACGCGTGCATGCT
243529
243539 GAAGNAGCCATGAAGCTTTCGGCAGGGGCGGAAAGTTCTGTAC
...
243796 GATTCTTGCCATGCCCTGTGTCATGGAAGGACATTAACAGCATATGCGAGGCGCTTCA
243846
243856 CAGAGAGCATACCGGTAAACGATGTCAGATATGATTTTGTGATACCGGCTATCC
243875
243916 TGTGCTGCTCAGATTTGCGGATTAACGTCAGCGGGGTTATTAACGATGGGCTACAGAT
243939
243976 TATAGCTGACGCGCATGAATCAGTATTAATTTTAAGAGTACTTACCCCTTTTAAT
243986
244036 AATTTG
...
244068 AGAGTCCAGATATACAGGCTTCAGAAATATGTTTATGATACAGGCTTTTATGCTTTAA
244118
```

```
...
236496 TTACTCACTACTGTGTTCTTTATCAATAATGTTTATTAATTCATCAACATCTGCCCGCAGT |-----|
236556 GAATGTTGATGTGAGAAAAATTTTATATATTTGGAATGTCACGGC 236546
...
236775 ATAGTACACGATTTAAGAGATTATGTCAGAAAGCCCGGTGATGCTTCTGACGATGAT |-----|
236835 TAATGTGATATTTCTATTGTGATCTGAGACCATATGGCAACAGTTA 236825
...
236897 CTGTAAATTCCTATAAGGTACGTTATGATCACAGCAGCTCAGATCATGAGCGTTTCCA |-----|
236957 GTGCTAATCCCTGCAACGCTGATTTCAACTGAATTCATCTACGAA 236947
...
237027 CAAATGTGTGAATGTTTCATACGGAGCGGGGTGCGCATGACTCCAGACGCTGAG |-----|
237087 GTGATATCAACACCGTCATCTTCATTTAAAGCGCTCACACGTAAGTAAACATAT |-----|
237147 CTCACGCCCGCAGTGTGATGTGCGACTT 237120
...
237418 ATTACCCGCGCATTTGAGCGCAACAATCTGACACTTCTGACTCATCAATCAACAGCATGCG |-----|
237478 CCGAATAACGATTTGCTTTTACGCTCCCGAAATTCACGAAAAAGCCATTGTTCAATT |-----|
237538 AT 237484
...
237573 CCAAGGCCCTCTTTTGTGTTTTCATGATGAATGATGAATGAGAAATCGACAGCGCATAGG |-----|
237633 GAAAAATAAATAACACTAATGCGTGTGTTTCTATCATATGATTTTATATTATTTAT |-----|
237693 GTCTTTTCTCAACCATTTTTCAGCAGACAATTTACCGCGGAATACAAACGTGA 237657
237753 ACGACAGTTCAACCACTATGGGGATAAATGCGAGATCAATCCGACACTGCA 237752
...
237908 CGATGCAACCTTCCCGAGCGGTAGCTTTGGTAGTGTCTGTGGGCTGGCGTTTCATC |-----|
237958
```

```
237968 AGCTATTGACGAATGGCTTACTTTACATACTGACTGCGCAGATCA
...
238144 TTAATTGATTAACAGAAATTTCCAGCAATTCATTTGTTGAAGACAGCTAATTTGTCCTATTTC |-----|
238204 TAGCGTAATTTGCTGTATTAACCGAAACGTATTACAAAAAGCAAG 238194
...
238333 CAATCAATTAACAAGTACCTCTGACCTGTGTTGCAATGTCAGTATTACAGCGATCTCA |-----|
238393 ATTGCTGTCAGAATCAGTTGACGTATGCACTCCGGAATTTAC 238383
...
238800 CAAAAAATACACAGCACCTTTGGCTATGTTGSCAACATCCGCATTACAGCGATCTCA |-----|
238860 ATTGCTGTCAGAATCAGATTGACGTGTGCAAGACCTGGAATTTTAC 238850
...
238908 GGGCTGTGTAATCTTTACTCTCTGATTTCTGCTGATTCGCGTGAAGAAAGACGCTGAG |-----|
238968 GATAAAGTGTTTATGAAGCATTTTATCTCAGGATGAAGCCCTGG 238958
...
239031 CTGTTGAGAAATTACAGTACAGAAAAAGAAATATGGGAGCAAAAAAGAAAGCGCTGAA |-----|
239091 TCAAAATTTGATAAAGAGATTCTGCCGTTAAAGATTACAAGCAA 239081
...
239163 AATAACCTCTCTGTTCCGCGCAGATACGTCAACAATTTTATGAGACAGATAGAG |-----|
239223 GGAATGTTAAATATTTACTCCGATAGCAATCGCTTTTGTCTGTGTATCCAGTGAAGG |-----|
239283 GGGGTAATTTTTCACAGCGCGGCC 239251
...
239355 TTCAATGACAGGAATATCTCTCCGGAATTAATTTGAAGAGCCAGACATGAGATATCG |-----|
239415 GTGATGATTCAGCCTGATGTTTACACAAAGTTTTTGTACGGAAAAAAGAAATTTGTG 239405
239475 AAAACGTCAGGACATCATGCAAGTTTTTGTATGTGCACCAACATCAACGCGAGGACAA |-----|
239535 AGGATAATGATCAATGAGCTTATTGATGAAGAAGTCTGCCGATGAGTGAGAACGATTCG |-----|
239584
239595 CTTCACTTTTCTCCAGGCTGCGAAGTCTGAGAGGATTAATTAACAT |-----|
239587
```

```
231553 ATGACGAAAGCCAGA
...
231570 AATTAAGGCGCAATGCTGGCGGTTAAAGCATCAACCGGAACACTGCAGACGATGATT |-----|
231630 ACGGAATCAGGCATTGGCGTGATATTAAGTGTGATTAACACGCCGC 231620
...
231751 CGGTTAACTCGGTGTGAGCGGTGCCCTGGCACACGCCGTTTGATGGAAGACGGCGTTTCAGG |-----|
231811 CAATGCGGATTAATCTGCCGACTGATATGCTTCCGCCAGAGCATCAGGTTCTGATGA |-----|
231871 ACGTAACAGCAAAAAGCGAGTGCCTCGTCATTAATATCAAGAAATCTCTCATTAACAT |-----|
231931 TAACTCATACGGTGAA 231891
...
231988 CGGTAGCTTTCTGTAATTAAGTAATAAGCCCTATTGGGGAATATGTTCAACGATTTGC |-----|
232048 CTGGCGTGATCAGCAACGAGTTATGCATTTGCTGAAGCATTTAG 232038
...
232197 GCTTTTANTGMAAAATGCAGGATTTACTTTTCATCAGAGGATATGTTGCTGCTTAAGTG |-----|
232257 GGATTTGCGCCCTGGTGTATATAATTTTATTTTGTGATCTCAAGC 232247
...
232364 AACCAAAATAGAAATTAATTTCTTACGCAACCTGTGTGCAGATTTCTGTGCTTAATC |-----|
232424 ATGTTGAGTGCANTGAGTTGGGGGCTATACGACGCTTACCCCTGTTATATGCGCTGTTTC 232455
232484 CTGATTTGCTTACTTGGGCCGTTGGTTATCTTTCTGCTGTTTCGATGATCCGACACAGT |-----|
232544 TATGTCAAGCAGGTGCTATTTTCACTAGCAGAAAAAGTAAATATPAAAAAGTCTGACTGC |-----|
232604 TATGTGCT 232556
...
232652 ATCAGCCCGTTGCGTAACAGTGAATCTTTTGTGATAGAGGGCAGTGGTTAACGTTTAAA |-----|
232712 TCGATAATTCATTCCTCATTTCTTTGTGGCGTAGTGTGGCATTAATCTGTTTTTCTG |-----|
232772 CGCTTCTCCAGCGGTACGTTTTTTTGGGCAAGC 232750

...
232892 AATATGCTGCTTATTTGAAGTGAATGTGATTAAGCTGTGTGCTGCTATTTGGCAACGCTTGTG |-----|
232952 GAATGGCGGATTTGGTTTGAAGCCTATTTTTTACTCTCTATGTACCGTGTCTTAATTTTAC 232942
233012 TATTTTTTCTATTTGTGATTTCTCTGCGCATACGATTTTATGATGTCATGTACATGTAT |-----|
233072 TTCCGTTGGGGCATTTTAATAAGTG 233042
...
233700 TTGCTGCAAGTGATGATCCTCTGGAGATTTAAAGGCTGAATCTGAGAAAGTCGCTATCC |-----|
233760 CAACATTTGCCGATGCAGCCAGCGCTGTACATGACAGAACTGTCTCTGATGGGAAATC 233750
233820 CAAGCATGTAAAGCAGTGTATTATCGACGCTTGAGAAATTAAGCGTTCTCAACTGGAG |-----|
233880 CAAAAGCGCTGGATTCGATTAAGGCTGGCGAGCTGGCAGAAACACTGGCTCCAGTCTGT |-----|
234074 CGTGATGAACACCAACCCGCAATGCCCTGGAGGCGATTACCGCTTTTGTGGCGCACGT |-----|
234134 GTGATTAACCATGAACCTTAATAAGTTACCGCGCAGCTGTATTAA 234124
...
234491 TTCAGGGACTGTGTAGCGAAGCAGGGGTATTCGCGGATCTGGCGGAAGAGCGCTCGCT |-----|
234551 CATACGCTGAAAAAATAAGTTGAGGCGGCATATCATCTGATCATC 234541
...
234721 TAATTAAGAAACCGCTTTTATGTCAAAATGTATTGAGCTGAATTAATTCAGATCGTCAGGA |-----|
234781 AAGCTATTTTGAATTAACAGGGGATTAATAAATCCCTTTTAA 234771
...
235584 ACATTAATGCAATGAAGATCTTTTAAAGATTTTCTATGTTTGTATATGTCAGATCGTCAGG |-----|
235644 TTCAATGTATTAATGTACCCCAAGACAAATGTTTCAATCTCAACGACACAAATTTATTT |-----|
235704 GTTTTTTGTTCAGGATAAAGATAAATATCAGAAATACTAT 235688
...
236285 CTGAGTGAAGAAAGATGAATATGATAGTAATAATGCCATCTCATTAATCCGGGCGATATGT |-----|
236345 GTTCCTGATGTGAGTGGAGAAATCAATGTTGACAAATGTCAATTAATA 236335
```

```
...
229020 TGCCGTGCGCTGCTACGTGCTGCGCTTCAGCGCGGACACAGATTAATAGGCCCTGGAG |-----|
229080 ATAATGCTGCGCGCTGACGAATTCGTTCCGCGAATAAACCGTTTC 229070
...
229286 AATCATGCTTTCGACTGCTAGCATTTATGTTTTACACGCCGGAGACGTCGACGCTCTAT |-----|
229346 GCCATTTTCACGGCAACAAAGATTAAATGGGCAATCTCTAATGGCGATGTCATCTTCAC |-----|
229406 TAACTTGAATATGCTTATGACTGACCCCTGATGGGGAAAGTCACGTGCTGAAAAACA 229389
229466 GTGTTACAGCTGGCGCTGGTGGCTAAATCGCGGATATCACCCTGGCCGAAGTTATCTA |-----|
229526 TTACGTTGAAAAAGACGCTGATATATCGCCCGTGGAGGCGGCTATTTATTCGACGATGA |-----|
229586 CCACTTAATAAAGAGCTG
...
229654 TAACCTTACCAGATGTTCTGTCAGCGCAATACTATTCTGATGAGCTACAGCGATGTC |-----|
229714 GTCTGAATGCGTGC CGGAGTTTCATTTCACCAAAAGATATATTCAGCGAGGTCCGAAT |-----|
229774 GAA 229721
...
229818 CGTTGACGCTAGACGCCGATTCCTGTTCAGGTGACGAGTGGACAAATAATGCGATGTGA |-----|
229878 CGCTTTACGAGCAAGTGTCTTATGTCAATAAAGAGCTTCGGCCTGGCAGATTTC 229868 229876
...
230085 CGGATGATAGCCGACGCGCTGCGTTATTAACGCCCTCAGCTGTGCTATATAGCGCTCGG |-----|
230145 ACAGCGTTCAATTTGAAGCAGCAAAATTTGCGCGCGCGTCAGAAATGTCTCTTTTGGG |-----|
230147 230147
...
230237 GGATTCAGTGCAGCGCTTTGCACAGATCCAAAAATGACATCGGCGTATCAGGCGCTCATCG |-----|
230297 TGGGGCGCAAAAATGAGCGTGATACACGCGATTATGCGTTTTCGGCGAGTGTCAATACCG 230287
230357 AAGTTAGCCACCATGACAAACGCTATGTACGCTCTGCAACTGTGTGGCTAATCAGCAG |-----|
```

```
226258 CGATTAACAGTTGGTATACACTTGTGACCTTGAATACCTTGAAGGATTAACGACAGTC |-----|
226318 GTATGCAATCTACTATTAAATCGTGCCTCCGCTATTTCGGTTAGCAATATTTTATAGTGTT |-----|
226378 ATGGCGCTGGGCGAGGAAAGTTGATGCGCTTAAAGCATTAATCTTGATATCAATTA |-----|
226603 GTTTATATTTTCAAGTTTTTTAAACCTTATGCGGTAATTAATGTATTCGATTAACGTTATTT |-----|
226663 TCCCACTGGTGTATATGTCAGTTCAGTAAAGAAAGAACCACTGACGAGCGCTGCATTT |-----|
226723 ATCTCGATAGTGTGATTAACGGGTTTGGGATCGTAAACCTGGGCAACTATCGGGG |-----|
226783 GTCAACAACAGCGGTGCGCATCGCCAGGGCAGTGGCCACGACGATGTGTGATTTG |-----|
226843 CCGATGAACCGGAGGAATTTAGACTGTGCAACCGCGGAGGCTATTTTGGACCTGTGCT |-----|
226903 TGACGATTAAATCAGCAGACCGGACACACATTTGTTATTTCCAGCAGCTCCAGCAAT |-----|
227054 TGTACCGCGGCTTTGTGGCTCCGAAGATTTACGCGGCAACCCGACGCTGTTAACCGAG |-----|
227114 TTCTGCAACAGGGGCTCCCGGCGCAGCCTGCCGCGCTGGGACGCTGACCGAGCGGTTAA |-----|
227174 AGAAGAGTATGAAAGCCAGCACAATGTGTGGGCGCTGTTTTCAGCCTATGCTTAC |-----|
227234 TGCCTCAAGCGGAGGTTATGCGGCACAAACGATTTTATCCATGCTCAGAAATACGCCA |-----|
227294 AAAGCGGCTTTTCTGGCTGATGA |-----|
227349 CGTACGGTATTTACGCGCCGCTAGATGCTATTTTACCCGACAGCTACAGGCGCTGGCT |-----|
227409 GTGACAGTCAAAAGACACCGAACACATGCTGCGTACCCGGGAGATATGACGACAAAGAT |-----|
227469 CCGCGACATATTTCTGCGGATGCGGTATATCGCGCATTAAGTCACTGCAAGGACACCAACCG |-----|
227529 CGGCAATGCGTGTGTGGCGAAATTAAGGGCAAAACGCGCGCTTGGCAGTGGCTGAC |-----|
227589 GCAGATTTTATATGTTGTACCAAGATGGATAGTGAAGATTAACCC |-----|
227739 GTTTTATTCAGAGATGCAAAATTTGCAATTTTGAATCTCTTTCGCCACAGAGCTGCTC |-----|
227799 TTTTCTACCATTAACCGCTATGTATTTGGGTGGCGGCAATTTTATATGGCGGATTT |-----|
227859 TATGTATTAATCTTTCTGTGATCCTGAAAGAGCAGACGATTCGTACCAATATTTGTCATGT |-----|
227919 ACAAAATTTCAATCAACACTATTTTGAACGTCGAAATAAATAATGAGCGCTGATTTGCCGA |-----|
227979 TTATCGCGGCTTAAGAAAGCTATTTTGTGCGACCCGACGCTGGGGGCGGATATTTCCAC |-----|
228039 GCTCTCCGGGCACTGAGTTACTGGCGTAA |-----|
228122 AGCCGTTACCCGCTTTAAAGCTGGGTTGCTTCGATTAAGATTTATCCGGGGAGCGATCTTT |-----|
228182 CCCGGGTAAGAGTGATGAATCAACCTGCGGAGCGGCTGTGGCGAAACGTTGAATGGCA |-----|
228242 AATATGACGACTGGCTGATGTGATGATGATCAACACGCGGCGGCGGAGGCGGCAATTTGT |-----|
228302 CGCTAAAAATGCGTGGCATTTTGTGATCAGGCATTAAAGATTATGACGATGCGCGATGA |-----|
228362 AGATCCCGCTGAGACCCGCGAGCGCATGATGATGGTACGATGGCGGTGACAAAGTGTCTTA |-----|
228422 TTCTGCTTAAAGAGAGATGACACCGCGGCGGTTTACGCGCAAAAGCTGCTCATTTATCGAAC |-----|
228482 GCAATCATTTGCGCGTGGTGG |-----|
228763 GGGGCGGTATGCGCTGATTAATTAATCTTATGTGATTTGCGATCCGCGCATCCGCGGA |-----|
228823 CAAAACATCGGCTATACCGCTTTATATAAAACAGTAACCCGGAAT |-----|
```

```
223883 CCGGCGGTGCGGTAGTGAAGCTTATACGCTTTGGTTTGTGCGCTTCGCCCTGTG 225207 |-----| |-----| |-----|
223907 223907 225925 225271 |-----| |-----| |-----|
223943 TTGCTTAACCTTCGCGGTGATTACTGACTGATTAAGAA 225318 |-----| |-----| |-----|
... 225318 |-----| |-----| |-----|
224253 ACCTTCACGAGGAAGCCATTAAATGCGGTTTAAAGGTAAATCCGTAGTGGCGATGG 225388 |-----| |-----| |-----|
224303 224303 225306 |-----| |-----| |-----|
224313 AAACAGGTGACGATAGCGCATTTGACCTGCTCTCTCAGTTGATCGACGAAGCCCTGAA 225394 |-----| |-----| |-----|
224322 224322 225404 |-----| |-----| |-----|
224373 CAGGCAATGCTGATTAAGCAATTTGCCGCGGCGCAACGTTTCGGGTTTATCTGACGGT 225454 |-----| |-----| |-----|
224409 224409 225507 |-----| |-----| |-----|
224433 ATTGGCGCGGTATCGACGGGCTGGATTACAATCTTTCTAATTAACAGCATGTAGAA 225514 |-----| |-----| |-----|
224481 224481 225524 |-----| |-----| |-----|
224493 GATATCAACTTAACGCAATCCATCACAATCTGCATGCCCTCAAGATGTCGACGACACG 225574 |-----| |-----| |-----|
224553 ATTTCAGGCATCTTGCGCGCAATATATCGCTTGACATTTTGGCCCTTAACATGAATGCG 225634 |-----| |-----| |-----|
224578 224578 22580 |-----| |-----| |-----|
224613 ACCAGTAACCTGCTACTGACTGCGGTGCATCTGGCGACGACGAGGCAATTGACGAGCGGG 225675 |-----| |-----| |-----|
224645 224645 225725 |-----| |-----| |-----|
224673 ATTGATCTCGGATTTGTGCTGAACGT 225735 |-----| |-----| |-----|
... 225735 |-----| |-----| |-----|
224745 ATGCTGACAGCAGAACAGATGACGCGGTTTGGTGAACAGTAAGTGTGCGCTTCGCA 225795 |-----| |-----| |-----|
224795 224795 225855 |-----| |-----| |-----|
224805 GAAGGTTTAAAGCGGCTCTACTGSAAGCGGCTCATCAACCGCGCGGCTCAACAAAGT 225885 |-----| |-----| |-----|
224832 224832 225895 |-----| |-----| |-----|
224865 GAGGCGGTAAAGGTTGCAAACGACTTACACCCAAATCAGCGCGCGCGCAGCAACGATGCC 225955 |-----| |-----| |-----|
224882 224882 225945 |-----| |-----| |-----|
224925 TCTTGCTTACACTAACGTGACAAAGGTATGACAGCGCGCATGAA 226015 |-----| |-----| |-----|
226025 |-----| |-----| |-----|
... 226015 |-----| |-----| |-----|
224986 CGCTGATGATCTTGACGCCATTCTTCCACACGGAACGGTTCCGAGTAGGCGATTAACG 226075 |-----| |-----| |-----|
225036 225036 226075 |-----| |-----| |-----|
225046 CGGAGGCCAAAGCCATCGCCATGTTGCGAGGGGAGCAGATCCCGGTTCTCGGCTATA 226092 |-----| |-----| |-----|
225060 225060 226142 |-----| |-----| |-----|
225106 AAGGCAAGATCGGCTATACCGCAACGGGA 226152 |-----| |-----| |-----|
... 226152 |-----| |-----| |-----|
225154 TCATTGGCACCATTTGTTGAGCATCATCAACTGATCGACACTGTGGCAAGCATCA 226212 |-----| |-----| |-----|
225204 225204 226186 |-----| |-----| |-----|
... 226186 |-----| |-----| |-----|
```



```
...
221628 TGGCGGAAAATATTCAGGATTTCGGGAAGCATTTATTAATAAAAAATGCGCTTCAAG
      |-----|
221688 ATAGCGAAGCTTTGCAAAAATGTTGAATAATGCGGTGCGCAGAGTGTGCATGACA
      |-----|
      221694
221748 TTGAATACTCCGATCTGAACGAAGATGTTATCATTCACTGGATAACGACGCTGTGGG
      |-----|
      221797
221808 CCTACAAAGTGGGTAAGATGGCGTGAATCAGGCCGCGCTGATGATAGCCGTGACACCC
221868 TCAAAAGAAAACAGGATGATCGTGGGTGTTCCGCCGGGAACAGAGCATTTTGGCGC
      |-----|
      221885
221928 TGTTGGAACAGCG
...
221952 TCTGCTGCGTAACGACACTTTTCCGGCGCATTCCTCATCTGTGTTCACGCTTCA
      |-----|
      222002 222010
222012 CGCTGCTGGGCTTACAGGGGTGGGTTGAATGTTGGCAGCCGCTGTACCCGCCAGCCCA
      |-----|
      222015 222046 222071
      -----|
222072 ACGCCGTGGGATGGCTTTGACTACATCCAGAAATGTAACAAACCATGCT
      222076
...
222195 CAGACACTTGTACGCCGCTTCGCGTCAATCCGATATGTCATGTTGTAAGCGCTGGCG
      |-----|
      222245 222251
222255 CTATCGTGTGAAGAGTATGAGCAGCGCTGATGACAGCGGGCGACATTTATGGTGAGA
      |-----|
      222256 222295
222315 TCTCTCTTACGCCACTTCTGCGATGCTTCCATG
...
222437 GATTACGTCACAGCGGCAAGGTACGGGTACCGAAGCAATGACCGTATTGAACGCTGCA
      |-----|
      222487
222497 ATGAAAAAGGTCTTGGCTAACCAAGATAAACTGCTGTGAGCTCCA
...
222557 GGCCACAACATTGGCGCAGCGGGATTGTGAACTGATTGCTGTGGCAACGCTTCCG
      |-----|
      222607
222617 GATAACCGCGTACTGCGAGCGCTGAATTTCAAGCATGACAGCCCGAATGGCATCTGAC
      |-----|
      222634
222677 TATGTGCCGAATG
...
222777 CCGCTGCATTCGGTCACTACCTATGATTCGAAGCGTGTGCTATCTCGGCGTGGTG
      |-----|
      222827
222837 CAGTATCCGCAATAGGCCACACGCTCAACAGCTTGTGAAAAACATCTGGCGCAGACC
222897 ACCACGTTGAACTGGGTAGCTAACCTTCCCGACGATACCTGGAAGAAGCAAGAGC
      |-----|
      222929
222957 TGCTGGCGCTGCTGATGAATCAACCA
...
223007 GAGTACAGCTTCAGGAAACCTACGCTCCGGAAACGGAAGCAACTTTAAAGCTCCAG
      |-----|
      223057
223067 GTTCCGGGCTGGAGCCGCGTAACACCTGCGCTTTCGATCAGCGTAAAGCGACCGT
      |-----|
      223098 223117
223127 GCGGCGACGTTTGCCCTGATGCTCACTACCGAAGCGCTGGAACAGCGCAAGCGTAATC
      |-----|
      223159
223187 AAAGCGAGCGGATGAGTTGGGCATGATCATTGGGATGTCCCGCGCCGAGAAAC
      |-----|
      223196
223247 AACTA
...
223283 CGCAAAATACGTACATCGGAATTTCTGCGTGTGATGAAATGCCATCCAACTTCG
      |-----|
      223333
223343 GGGATCTCTGAAGGATCAAAAGGTATACCACTACGCTTGGCAGAGGGAAAACTGGCG
      |-----|
      223399
223403 CTGGCGCGCTGACCTACGTTATGAATTCGTCGCGGAGAACTCCACCGAGTATG
      |-----|
      223404 223430
223463 GTTGGCGCGGATGATGACTTCCCGTCAATGCTGTATATGATGCGGTGACACA
      |-----|
      223495
223523 AAGATCCACATGACGGCTGAGCGGAGGACTATACGATTTACGGTAAGATCCGAAAGT
      |-----|
      223547
223583 TATGTTCCGGGCAAGAGGCTTGCATGCTGCTGGAAGATCCGACAGTGGCTGTGGCG
      |-----|
      223646
223643 CGTGGCGCTGAGGTTCTTCGGGAAGTGTGGGCTAACGCAAGTCATGACGAATACCTAC
      |-----|
      223646
223703 TTGACGCTCATTTGTTGAAGAAAAATCCGCTGATGATGGCCATTGAACGGCGC
      |-----|
      223706 223741
223763 CTTACGATGGCGGCACTCTGCTGGCGATTTGATTTGATGTGATACAGCAACGCT
      |-----|
      223787
223823 AGCGATGACCATTTCTGATCGAATGACCCGATTTACAACACCTTCCCGCGCAAAAC
      |-----|
      223823
```

```
...
218612 ATACTCCCTGAGTATTTCGCATTGCCGTACCCACGCCATTGCCAGCTGTCAATCGTTTATT |-----|
218672 AACTCAACATATGTGATGAATATGTACAAATATATATGATTTAAATG 218662
...
218769 GCGAAAAATATCTGAGCAGACAACCTGCGATGTTTATGCCCTCTATTAATCAACGACGTGC |-----|
218829 AATTACGGCGCTGAAGACCAAGCCATCTCGACGGCGATTAACCTGCATATTGTGCAGTGC |-----|
218836 218861
218889 ATCTTGCCCTGCCCCAGAGCGGTGAGCGCATTTGTGGAACAGATAGAGGGCCCAAGCTGGCA |-----|
218934
218949 AGATTGATGTGCTGTAAATTAACGGGGGCGCATCGTTAAAGACAGCCTGTTGCCCTGATGA |-----|
218978
219009 GTTATGAGGATTTCACCCAGGTAT 218978
...
219066 CGAAGAGCGGCTGATGCTGCTGTGCGGCGAGAAATCCGCCATTATTACGTTGCT |-----|
219116
219126 CTATTGCGCGCTGATCCCAAGCGGTGCGTCAAGGGAATTACAGCGCGTCAAAAGGGCGGA |-----|
219133 219146 219181
219186 |-----|
219186 TCCTAAGGCTTCAACCCGACGCGTGGCAGCAGAAATGGCACCCCTGGGGCGGTGC
...
219318 GTACCATCCGCTTACGCCGCTGGGTAATGTGAAGAAGTGGCGAACCACATCGTCTTCT |-----|
219368
219378 TAAGTTCCCTGCGCTCCAGCTACATTTGCGGCCAGACCATGCTTATCGATGGCGGCTG |-----|
219416
219438 TCATGCGGTGATTTATGTCTAATAACAGATTCCTTCGAAAATCTTCTCGAAAATGGCG |-----|
219498 |-----|
219501 GCTGGCGTCAAGCGCGCTGCTGATGGTGCATAAATTTGACGATTATAAATAGCGTGAAT
...
219667 GGCCAGGAGATATCTCTTCTCTGACCGACATTTGCGATATCTGGCCCGCAACGTTTGA |-----|
219717
219727 GGAAGAGCTGAAGTGTGTACGTGATATTCAACGACGATTCACCGCCTGAATGTGGA |-----|
219739
219787 GATTATCC
...

219804 GCAGCAGGTACGGGCGTGTGGCGGCTCAGAAACCTGAAATTTAAAGATTACGCGTATC |-----|
219854
219864 CGGGTGAATTCTATTGATGTGTGCAGCAAAATTGGCTTTTTCGATGC |-----|
...
219987 GTGAACCAAGTAATTCATATAAGACGTAAATATTATGTCTACCATGATGATACGCTATTG |-----|
220037
220047 AACAGCGTAAGAAGTTCTGATGACCATTTAAACTGAGATCATCCAGCGCATATATTC |-----|
220092
220107 AACGGATGAAGCGCAAAATGATGATGATCTGCGCTGTTGGTAATGCTTTTAAACTGC |-----|
220167 ACTCCGTGATGCAACGGAATTTGCGTTGCTGTGATGAAGTATTTGATATCAAGATCA |-----|
220190
220227 GCGAAGTGAAGACCGGCTCTATATGCGCAGTATCAATATGTTGCGCACTTCGTTATTC |-----|
220235
220287 AAAAGAAAACGATTTAACCTACTTATTAAAGATTATATATCAGACACGTTTATGAATC |-----|
220295
220347 TCTT
...
220445 CCGTTGCGTGAAGCGCGTGTGTGACATATTCATCACCCTCTATTGTTAGCGTCATG |-----|
220495 220504
220505 |-----|
220509 220518 GCGATGATGCGCTGGCGGCTGATCAACCAAGCTGTGTTCCGCAATGTCTCCATTCGTG
220565 ATGAGCAGCAATCTACTCACTGTGTTGATGAAGAAGGCGCACCATTCGCGGAGCGTCT |-----|
220625 ATGTGCTGTGCGAGATTGACGGTTATTAACCTGCTTTCAGAAATATTGCGGCTCCGAT |-----|
220685 |-----|
220687 TGATCGCAGCATGAATACCAATTCCTCGAAAAAGCGGAAGAACTGATATTCATTCGAT
...
221186 ATAAACTCATGGGATGATGCGCACAGGAAGAATCGAAGGATTTGAGGCTGCGGATCGG |-----|
221236
221246 TGCTGTGGAAGTCAAGACGTGGGCTATGTGCTGAAGGCACTCTTTTCCCGCAGCGTTGC |-----|
221298
221306 |-----|
221306 AGCGTTTCATTGCGCTGACCTGCTGGAAGGATTACGCTGTGACATCAGC
...
221455 GAGCACAGTTATATCGACGCGTCAAAAAATTAAGCGCAGCTAATGAATATGCGCGCT |-----|
221505
221515 GACTATAGCTTACGCGCTTAATTAACCCCAATTGGCGGATTC
```

215809 AAGAGGTGTCACGCTTTCTGCTCAGAACGCTTACACATATACAGCAGCATCATCGCTG
215835
215869 AAAATGTCGCGTTAATCTGCTGTCCAGGGAGACGCTGGATTAATACCCGCGCTTCTTA
215900
215929 GCAGTGGGCGTGATGCCATCATCCGTGCGGAGGAGCGTTCTACAAACAATATGCCACCA
215963
215989 CGTACAGGCGCGGTAACTCTGACCGTTATGCGGCGCTGTAACACAGCGCATGCTG
215994 216009 216020 216033 216046
216049 GCCTGGAGACATATACCGCCACGCGCGTGTGCGCTGTGACAAAAACCTGGATCTGACG
216051
216109 TTGATTAACAGTGTCACTAATATGTTGATCAGCGGTAAAGAGATGTGCA
216110
216409 TTAACCCCGGTAAATTTATTAATCTCAATATGCTGATGTTATGCTGCGTGTGCGG
216459
216469 TATCGCAATTAAGTTACGATAGCGGTAAAGACGCTGTTCTGCGGCTTCTACGCTT
216501
216529 TAGCGTTAGAGCAAAACGAACTGATACACCGGTACTATGTGCGCATGTAAAGTG
216530
216895 GAACGTGAAGATTAACCTTTCTGGGTGTCAGTTGATTTCTGAAGTGCAAAAAGCCATTGCC
216945
216955 GGACGTTGTGTCGCTAGTGTGTGCAAAAAATTCAGCTGCTTGCAGGATTAT
216956
217063 ATTTCTAAAGCAGTACGTTTCCAGGTTAATGCCGGAAGATGAAAAAAGCATTTGTC
217113
217123 GAAGTAATGATATCACTGATGATGATGTACCTTTTCAGCGTAAAGATATTCATATAT
217165
217183 CGTGATATTGAACAGGGGCTGGAAAAATCTTCAGCGAGTACCTGACGTTGATGTGAAAAAT
217225
217243 AATTAACCCGGGACACGSGATGTTACAGCATGCTTCATATTTGATACACACCGGGGG
217276
217303 AAAAACTGGAGCGTTGCGGCAAGTTACACAAATTTGGGCGAAGAAACGGCGCTTAC
217312 217339

217363 CAGACACCTCGGTGGTTATCTGTTTAATCC
217406 GGGATCTGTTTATCTCGCCGCTACGCTAGTACTACCGGACAGTATGAAACGTCAGTA
217456
217466 GTTATTAATCCATTCGCGTGGGTACTGGAGTACGCTTTTCTA
217546 TCGTACATCTCGCTGGATTAATGTTGGCATCAGCGAATCTTGAGCGCAAAAGCGACGAT
217596
217606 ACCGTTATCGGATAAAAAAGTTCCCGGTAGTGCAGAAC
217740 TTATTAAGCAAGTTACCCGCGCTCTGGGATTCACACCTCTGTTGCAACGCTTTT
217790
217800 AACCTGTTGCGTGTGTAAGAAAAAGCCGATATGCAATATGCGACGCTTACCCGCTAAG
217821 217840 217843
217860 CAATATCTTAATTTTGAAGGTAACTATACCCGCAAAAT
217928 TTGCCAGTACGACACGCGGTGTAAGTACTGTTTCAAGATAAATTAAGTTGGCATGTT
217978 217982
217988 GGAGTAATTCGCGGCTTGAATAATAGCGTAGGTTTGTACAGGATGATGCT
218075 GAATGAAGCTAATTAATTAATCCGCTGTGATTTTGGCAAGTGTATCAGAGCGCTTCT
218125
218135 ATGGCATGAATCACTGTGGGCTCGGCAGTTGGTATGATGAACATTAAAAATCTTGTG
218137
218195 AGTAACTTTTCCGTTTCTAACCCATTAAGTATCCGGGACATTTAGATATTGATCGGG
218232
218255 TGAATGTAAATTTTAATTTTTCGTACCAATGT
218352 CATTATTAAATGAGAGATGCTTTTATGACGCTGTTTAAATCTTCATGAGCGCTTACT
218402
218412 CATTAACTTCAGAACCTGATTAATGAACGGGCTCTGATTTCTGGGCGGCAAAAGAC
218469
218472 GTTAGTTAAAGCATTTGTTGTGTTTTCGTGATGCGGTACGTTTCCAGATATTTGAAT
218474 218481
218532 AAAGC

```
213315 |-----| CTGATATACGCGCACGCCGACAGAAATTCCTGCCGTAACGTGTAGTTAAAGCGCAAGTG  
213321  
213375 |-----| TCTGCCGATGAGCTGAACGTTGTGGTGGCAATTAATGTAAATGCCGACGCGAGGTG  
213390  
213435 |-----| ACCGGTAGCGT  
...  
213511 |-----| ACAAAATCAGTCTGTTGTCAGCACCGAGAAAGGTGTGGGGTTCCGCAACCTCGCGCTTAATG  
213561  
213571 |-----| CTGGGGGTGTTAATGTGTGACATCGATTCCAAAGGTAACCTGTTAAACAGTAAAGCCC  
213624  
213631 |-----| AGATTAGTCTGCAAGCACGATCAACCTGACAAACAATGGTACTCTGGA  
...  
213727 |-----| CCAACAAGAAATCTATCTGTAATACCCGTGCGGTAACTCTACGATGGCGCATATCT  
213777  
213787 |-----| ACCTTAACAGCGGTACGATTGACAACTAAGTAAAGGCAAGCTTGGCGGTGCAAGAAATGCTGG  
213847 |-----| CGGTGTATACCAATTAACGCCGCTGATTAACCTGTGTAAGGAGTTCTGTGCGGATG  
213861  
213907 |-----| AAGCGGGCTGCTGGCGCTGAAAAACGGAACGCTCAACAACAGCAATGTCAAGTTGCGG  
213920 213935  
213967 |-----| GTGGCTATGTGGGTCTTGAAATCCGCTGCGCTGAATTAACAACAACGTTGATATCCAGACA  
|-----|  
214027 |-----| CCGGCGATATCGCATTTATTCAGTAACGCTAATGTGATTAACAACAAGTCTGTCCGTT  
214029 214035  
214087 |-----| CGTCCACCGGGCATATATCGTTATTTGGCGCGGACAGTAGCTAAATATGTTCAACCAAAA  
214101  
214147 |-----| CCGCGCATAC  
...  
214204 |-----| TTGGTGGACAACATCAATTAACAACGCGGACAGATTGGGTCTAATGGAAGATGCC  
214254  
214264 |-----| TGTCAAGTTACAGCAGCATGACGACTATGCGGGCAAAATTCGTCCACAGCAAAATGA  
214284  
214324 |-----| TTATCAAGGGAAGCTCTCTGCGTAACGATACCGGGGGGATCAGCGGTAACAGGATTTG  
214347  
214384 |-----| AAGTCGCGCTTGGCGGACGCTGACCAATAATATTTGGCGTATCAGCTCTGAAGAGGG  
214386  
...  
214474 |-----| ACCAGCGCGCTTCATGATGGGGCAACAATCAAGATGAGTGATGTCTGGCGTCATA  
214534 |-----| ACAACACAGCGCTGATGCTGGCCAGCAAAAACTGAAGATAATGCCGCGCAGTATCG  
214541  
214594 |-----| AAA  
...  
214678 |-----| AAGGCATCAGCTTCCGGGCAAGACATCTATAACAACAACAGCCGCTTATGCTGAGG  
214728  
214738 |-----| ATGTCCTCTGACTCTGACGGCCAGAACACGTTGCAACACGGC  
...  
214966 |-----| CACTGGAAGTGTGAATAGCCTTAACCACTACGGCTGGATCAGCGGTAAGGGGATGTG  
215016  
215026 |-----| ATGTCACGGTGAATTAACGCCACCTGTATTAACCGCAATACCATTC  
...  
215089 |-----| ATATTGCCGCTTGAAACGTTATTGAAAACTGGAAGATATTCTGCTGGCGGACCTGA  
215139 215147  
215149 |-----| CGATTGAACACCATTCGCCATGTGACCAACAACACTCAACACCAATATGTGGGCGAGATA  
215152 215161  
215209 |-----| TTGTTATTAAAGCGGTTAACGATATCAACAACCGTGGCAACATTTGTCAGTACGCTGACC  
215226 215259  
215269 |-----| TGAAGTGACGACCAAGCAACCTGTATTAATCTCTATATGTAAGGTATGGGATA  
215276  
215329 |-----| TCGCATTTGTGGCAATAGCGTGGCGCAACAATAACGGACCATCGAAGCGACAGCGATC  
215375 215382  
215389 |-----| TGATTATCGATTCGAAGGGTAAACGTGGTAAACAACCGGTAATCTGCATGCGTTGAACG  
215449 |-----| GCGTGTGTCTGTTAAAGCAACAATCTGACACGATTAACGATTAAGGTAATTCGTTATG  
215481 215508  
215509 |-----| GCGATGACAGCTGGCAGTACGCGCAACTACGACAGCTATAAGGTTCCGTACCTCTG  
215513  
215569 |-----| AAAGCGGCGAGCTGACTCTGACGCGCAACATTTGTAGCAAGCGCTATGTTGATTGCCG  
215574 215607  
215629 |-----| GTGAGAATGTTCTGTGATGCTAATTCGACGATTTAACAACAACACTGCCGTGATCCCG  
215657  
215689 |-----| CGAATAAAGCTGTATTATTAAAGCTGCGGCGCAACCTCGAANAACCGGAGGGAATACT  
215709  
215749 |-----| TCCTGCGTAAATACGGCGCGCTGTTGAAATTACGACCAACGTTGGCGGCGATCTAGATA  
215787
```

```

...
209605 GTAAAGTCAAGTTTITGGGTCGCTACCTGATTTACCGACATCGACCTGGAGCAGTAGAA |-----|
209655 AATTTTGCATCTGATTTAGTCAGGGCCAGGATGAGCTGGTCTGACTT 209655
...
209757 AAACAGACTCGCGCATTCCTGGGGTAATTAATTATTAACAGTATGAGCAGGCGTAACGGCTTGAC |-----|
209817 CTGGCGCAGCTTCCATGCGAATATGTCTGGGGGGTGGGCATTAACGG 209807
...
209870 TTGTCAGAACACAGCTTCATCCTCAACATGATATTAATGATATAGGGCTGATGCGTGGTA |-----|
209920 TTTCCGCAACTGATGATGGTTTTTTTCTTTCCTTTCGCTCACCGAGGAAAAATA 209920
...
210159 TAGCTCAGGGGTGATATTCCTGCGACAGGGGTAATTTTACCGGCACTGGCAGCAGCGCTATCG |-----|
210219 ATAAAGAAAGTAAACGGCAGCAGGAAGTGTAGCTGCTTCGGTATTCGCGCACTTTCAGAGGTC |-----|
210279 AGACCCGGTACGTTAGTTCGGCGCCAGCGAGAGTTACTG 210260
...
210318 TTACCAAGAAAGTGTGCTGGAATCCATTAAGTGAATGCCAAAGCCTTTCGCCGATGCGTCTGCG |-----|
210378 GTATCAAGGTTTCCCATTAACACCCGTCGTAGCTCAACGTCGTAGCCGACGAAAGGTCGTG |-----|
210438 TTGATTTGCCCTGATCTGCACCAAGTTGAGAGCCTGSCAGTTAGACACAGAAACGTTAAACGGC |-----|
210498 -T- |-----|
210500 GTCGGGCTGCGCTCTGACCAATCAGCTAAAGTTGACACGGAAGTCAGCCATGGC 210495
...
210726 AAATTTAGAGTTAATATAGAACAGAGTTAATTATATGCAATTTGATACATCTGACGTTTATA |-----|
210786 AAACCGGTAATCTCCTTAACGTATAGTGCCTTGAATAATGTAAA 210776
...
211235 TAGGTCAATTTTTCATGCAATTTTGAAGACAAAAAACACTCAATTAATAAAAAACGATTAATG |-----|
211295 GTGACCTAGTTTTTGACCTAATATCCCTGAAGGTAAATATAGATCATCC 211285

```

...
211524 GTTCTTTATATTAGAGAGTACACGGGTATATCTGTATCTCTTAAGAAATGACGATGTC 1-----1
211574
211584 TGCAATCGATGAGCAACAATATATGACATTGATCAGATGATATACAA
...
211764 TGAACTATTAATCTTTTATATGTCACAAAAAGCCGTTTATTAAGAAATAAAMCAGCATATTT 1-----1
211814
211824 AAGAGTGTGTGGCAGCGGAGCAAAAAGGGTAAATACAGTCATTTAAGAT
...
212091 AGATGGGAATGAAAAGCTTGCCAAACCGTTTGGCTCTTCTTTCCCTTTCACAGATTAAGC 1-----1
212141
212151 CAGGCATTAATGTGGCTGCTTAAAGCCATTTGCGCAAGCCAAAGCT 212141
...
212267 ATAACTCTTAAGGGTTACATGGGTGATGCAAAATGTAATATATAAGAAATGTCGGCTGGC 1-----1
212317
212327 CGTAATGACCATTAAAGATCTAACCCGACGTAATAAGATATAGAA 212317
...
212450 TTATTTGGGATGTCTGAAATGATGTGTACAGAAAACCTACATATCATTTTCAGAGCTGCT 1-----1
212500
212510 AAATGTGAATTTTTTTTACCCCAAGAGCAAGGATGGGGACTTTTAA 212500
...
212712 AGTCATCAAAAAGTGATCTGCATGTCCAAAATGAATAGCATCAATAAGAAAACGTTAGTG 1-----1
212762
212772 GCGATGATTAATTTTAAAGTAAGAAAGCAACGGGTGGGAAAGCATTAACCCCTATGTG ---1
212776
...
212851 CTGTAGCTTCTCAAGCGGGGAGATTAGTCATTAATAAATGGAACGGTATATTAAGGCAATG 1-----1
212901
212911 GTGTGCCATGCTGATGACATCAACAACCTAACGGTAGCGGTTATCTCATTAATATCTGGG 1-----1
212919
212971 ATTAACCTTAAGCTGTAAAAAATGCTGTCTTTCATATATTAACGCTAATGATTCACAGTA 1-----1
212979 212997 213012
213031 CTTCACTTGC CGGAAATATTCAGGGAAAACAGTAACTT
...
213195 GTAACGGTGGCGGTTCAATCAATACAGTAACTTAACCTTAACCAACGGGACGGCGGAT 1-----1
213245
213255 ATTCAGGATGACAAAGCTGGCCGTTACTCGGTGACAGGGGGTACCATTAACGCTCGGTAA

206941 CCCCCGTACGGGCTATACCGGACGTGTAATTATTCTCATCCAGCAGCATCGGCTCCCATTT
206961 |-----|
207001 GGCAGTCCGAAACCATCTGCCATGGGGATTATATCAGCAACGCTTTACAGAGGAGGGA
207061 CGAATAATTAATTTCACCGCGCCACCCGGAACCGGTGCTACTGCTTTCACTGTGCAGC
207121 AACTGGCATCGCCATTCATTCTTGATGAGAGATATACGGTTTGAGGATAGGGC
207129 |-----| 207170 207178
207181 GTTACTGGCGGAACAAGCGCATAGCCACTGTCATTATTTCCAGCTGACTGAGATTATAG
207183 |-----|
207241 ACTTTGACCTTTTCGGCCTTTAGCTTCGACAGGCGGAACGTTTCAACCAGATTAAGGT
207301 CCGAAGTAATGCGCGCCACCATGCACAGCCATCGC |-----| 207280
207360 CCAGTAATCTGGCCCTTCGATGCGTTCATTCGACGAGTGTGTAATAACGCTTTTGG
207420 CATGTTCCGCTAAGCGTCTGCTGCTTAATTGTTGATCGCGCATGAGCTGAGCTGTA
207433 |-----| 207465
207480 GTTGGTGTCTGTGCTTCATCAAGCATTCGGTTAGCGAGC |-----|
207621 TGTACCGCTTCGATATCATAGAAATGTCTTTATAGCCGCCATTCTTTGGCGTCGAC
207681 GGAAGGTTCACTGATGATCCATGCGTAAGCTATTGCTGTAACCT |-----| 207671
207795 GCCGTAATTGCGAAGGCTCTGACTTAAGSTAAGATCGAAGCGCGCACTGTTGACGCGCACTG
207855 GCTGAGTCCCATGACTGTTATTGCTGGCAGCATGACGCTCTCCAGCACATCAGCCAG
207915 ATCAGCATAGCCGCTGTAGAGTATCATTA |-----| 207890
208058 GGTGGCCCCCAATGCCCCCAGCGAAGTGCATAGACTCCGCCAGCATCAGCGCTGTGT
208118 AGCCATCAGCGATAGCGGAACCACTGTAAATGTCAGGTAATAGTATCCCGTGTGCGC
208125 |-----|
208178 AGG |-----|
...

208245 GGAAGTCTCTGAGCAGCATGATTTGTAACGGCTGAAAAAGGAGACATGAACGACTGAC
208305 TCGCCGCTTAGCTTCGTAACGGTGACATCCAGATTCGCCGCTGATAGTGGGGATTAG
208365 |-----| 208338
208425 CATTAACGAGACTGTGGCTTTGTGGCGCCAGCCGGAATATGCGGCGCATACCGCG
208440 |-----|
208485 CATGAGTCCGGCAGCATACGTTTCATCGGTGCGGAGACTTAACGCCGTATTAACGATGTC
208545 AGAGAAAAATCTTCGGCTGTGTATTAAGCTGCCCATTAATGCTATTTATGCGGCG
208605 CAGCGGGGCTTGCATAATAGCTAAGATAATTGTTCCACTGATTCCTTTGTCAATATCCCA
208610 |-----|
208665 GGTCAATGTTGATTAACGATGACGATATTGCCAGGACACCAAGTTGATGCCACCAATTAAATGA
208725 TGCCCATAGCGAAGCGCTGATAGCATTTGCCCCGTAATAGCAAGTTTCAATTAATAGTT
208736 |-----| 208770
208785 GGCATTAATTAATCATGAACGCCAGCATGCTCCGGTATTCAGACTGCTGGGGGAACATA
208801 |-----| 208807
208845 GCCCGAAGGCGGACTTAATCAATTGACTCTGTGGACCGATAAGTCCAGGCGCAATTGTGA
208905 CAAC |-----|
208967 TTTCCCTTAATCTTCCCTGTTTGTAGGCGCTATGCGGCTCCGCACTTCAGGCGTACAGC
209027 AGGCGAAGCGGACCATCAGCGGCTCTTTAAGCTGATTCGA
209108 TTTTATTAATTTGCCCGGTGCAACAGACTCAGCTGTTGTGAGTTTGTCTAATGACGACTGAC
209168 TAAATCTGCTGCGCGGGAAGAGTTGCGGGTCAAAAGGTGAATGTC |-----| 209158
209351 AAAGAGAAAGAGTTATTAAGGATGTCAGACGTAACATTTAGCGCATTAAGCTGAGC
209411 GTCACTAAATGATCTGCGACCGGAGCGACACTGATGCCAGAGGTAATGTCCATTTCA
209437 |-----|
209471 GTTGTGCTTTGAGGAGCAATCA |-----|

```
204251 AAGAAAAGGAAATATTATAATGTCCTCCGATTTTCATTTCAATGCGGAGACGTTGGTC      |-----| |---
204311 GCCAGGATTCGCGGACTAAACATGCTTTAGCGGGGCTGACCAAGGCTATGGCGATGA      --| |-----|
204313      204323      |-----|
204371 TTTTGGGGAATATATATATTCAGATTAAATGATATGCCCCAGAAATATATGTAGTGAAT      |-----|
204431 GACCGCTATTATTATATAACCTAAATAAAGAGTTAATAAAGGCGCTATACGGC      |-----|
204491 TCAACGCTGGGGAAGGCGACAGATCTTATGCGGCAATGGTGTCTCCGACAGTGC      |-----|
204493      204803      |-----|
...
204557 TTATGTCATGAGACAACATATTAGTCATTTGATGGATATTCATTCGTTAACGATTTT      |-----|
204617 GCGTCGCTTGAGGTCTTTTGACTGGAATAAGAGTGGAGAGATTTTGTCAATTTCTCCACT      |-----|
204677 TCCGCTACTTCAGGTGGGTATAGTGTATTCGTTAGCGTTGAACCAATCGCATGAGTTT      |-----|
204737 ACATACCACTCAGACATATTATGGCTAATTTGTTGTTTATCTGTGACCACTTATTCATCG      |-----|
204797 AGATATGTCGCGGCAACACACACATACGTTATTTACTGTCTAACCAAGTACCGTGTGA      |-----|
204803      204823      |-----|
204857 CTGAAGGCTTTTATCATCATATCCCTTTCTTCTGTCGACCTTTGTAAGCCCCAATGCTGATG      |-----|
204877      204877      |-----|
204917 GAGGTGAAGAGATATTACAGTGTGTGGTTATTAAAGTCATATAGTACTGCTCCG      |-----|
204977 TCGTTACCCAGANGCCAATTTATTTACGCATACCTTCTTGGGCGCTCATAGTTACAGAT      |-----|
204981      205019      |-----|
205037 TTTGGCCTGTAAAGATTTTAAGTTGCGTATTTCAATTTGCCAGCGTTAAATGTGTGGATCT      |-----|
205097 TGAATGGGAGCTTCTGCGGCAATTTCAATAAGATT      |-----|
205076      205076      |-----|
...
205295 ACCAGTACCGGAAATGATCTGATTTGATGACAGATCAGTGTCTTTTCCGATCGCCATCG      |-----|
205345      205345      |-----|
205355 CGATAAATATATCCGTGAAGAAGCGGCAATGATCGCGCTTCCACTG      |-----|
...
205471 GCCCGCACTTTTCCCGACTGACTGTCTGATTTGGCAGTTTCTTCAACGCGGGGATAA      |-----|
205521      205521      |-----|
205531 GAGTAACTATAGCTGAGATTCGCGGGGTGAGTCGACCTTCTAATG      |-----|
...
205604 CCGCTGATTTCTCCGGGGGTTAACTTCCGGATTCGCCGTTGACCAACACGCTCATCG      |-----|
205654      205660      |-----|
205664 CCGTATCGGGGTGAGCACTGTTAGAAATGTAGATCCCAACGCCCTTNGCCATCTCTGCTG      |-----|
205665      205701      |-----|
205724 CGTCATATATATCAGAGACGAGCGGCTGACGCGCGCGCTGTCTATTTGATTAATCCCACTT      |-----|
205746      205752      |-----|
205784 TTTGCGCGCGGTATATATGACACCTTCAGACACCTGATTTCTTAATGCTGTCTGACTATTCG      |-----|
205844 TAAATCCCTGATTGAACGAGTTACTACTGACCTGACGCGGACATTAAATTTAC      |-----|
205844      205844      |-----|
...
205909 CCGGCATTCAGTTGTGCTTCTGCAATTTGTGCGGACAGCAATTAACGCGGTGCGCTCGG      |-----|
205969 GCAACGAGGTGGCATTGTTGTAGAGTCGATTTGCGCAACGCAATTC      |-----|
...
206295 GATTTTATGTTCCCTGAGTGGCATTAAGTGTGATGGGACATTTCTTCCAGTAACGACCAAG      |-----|
206345      206345      |-----|
206355 CGTCAGCGCGGCCCATTTGCTGCTTGTAGACCGACCAAAAAGCCAGACGAGTGCATTA      |-----|
206358      206358      |-----|
...
206410 ACATCACTCAGACCATCATGAGCGCGCTGCAATGTAAGAGCCGCCACGCGGTGCGGCA      |-----|
206460      206463      |-----|
206470 TTGCTAGCAGCAGAAAATAGATGTTGGCAATCTGTGTGCATGCATTTCCACACATAGC      |-----|
206475      206475      |-----|
206530 G      |-----|
...
206597 CTGCTGGCGCAAAATAGTGTGCGTCAATTAATTTTCCGAACCTCAATTAAGCGCTGTCC      |-----|
206647      206647      |-----|
206657 GACCATCTTCAGTTTGACTATCTATCTGTGTGTTGTAGAGTTAGCCCA      |-----|
...
206821 ATAGGCTATTTGCTGTCTTCCGCTGATATCAAAAGGCAATTTGGCAGTATATGTTAGCA      |-----|
206871      206871      |-----|
206881 CTTTCAACCCAGGAACTGACAAGTGCCCTTTTGTCTGTCTGTGCGGAGGTAAATTTGC      |-----|
```

```
199538 AGAATTTATTTTAAATGAACTGCAGTTCAATCAGCATCAGCCAAAAACGCTTAT 201593
199588 199588
199538 TGGCTCATGATATGATATGACACCTTCAAAAAAGTCAACGACACCTGGGGCATCCG 201593
199658 199661
199658 GTTGGCATCAGGTGATAAAAACAGTGTGATATCATTTGGGAAAAACATACGACAGAT 201593
199718 199751
199718 GACCTTTTACGGCGCGCTCGGGCGGAGAGTTTGGCGCTTGTGCTGACGACATCGATCT 201593
199778 199821 199832
199778 GAACGTGCGAAGCTTTAGCGGAAGGATTCGGGAAATGTTGAGCGCTTAACTGGCGAT 201593
199838 199867 199880
199838 AATCTGATATACGCTATATACCACAAAAAGTGCAGATTTGCGCTTGTGACTCAG 201593
199898 199867
199898 GGAATGCCGTTAAACCGAATGAGATCTATCGACTGCG 201593
199999 200049
199999 GAATTTTGTGAGTCAACATAAAGCGGATTTAGCGGATGATTTGGTAAGCGCTAAT 201593
200059 200049
200059 AGCTATTCGCAGACATGCGTAACAATTTTCGAAATATCATAGTAC 201593
200226 200276
200226 AAAAAAATGTTTGTACTCATTTATTAATACAGGCAATTCGGGGTGTACTGTGCTTGT 201593
200286 200276
200286 ATGTCAATTTCAATGAATGAGATGAGATACGACAATTACCGCCGCTA 201593
200488 200538
200488 ATAATTTCCAAATATGTTCTTTCAGTTAGCGGTTTCAAAAAAATGAATGTAAGGATTT 201593
200548 200538
200548 ACGGTACGAGCTTTCACAAATAAAATTAAGAAATCAGCATAGTATC 201593
200906 200956
200906 CATTTTCCATTATCTATCTTCTTCTTTTATCTGTACTCTTATGCGACGCGTGAT 201593
200966 200956
200966 AATCGGAAATGTCCCTGGCAGAGAAATTAAGAAAGATGTCA 201593
201030 201080
201030 CATTAACAACAGTATGTGACACAGACCAACGCTGTCTGGGGTGTGAGCGCTTAATG 201593
201090 201080
201090 CGTTGGCAACGCGAAGATGTAATAATATATCTCCGTAATTTTA 201593
201492 201542
201492 CTTAAATAGATAAAGCTTTATAGACCGCTCCTACAGCGGAGCAAGCAAGCGCTGTT 201593
201552 201542
201552 TTAGATATTATTATTAAGCTGAGTAATCGACTCAATTTGATTAAGATTTGCAAGAGTGTG 201593

201612 201593
201612 AGTACCTCTGCACGCGGGAATATCTTATTCATATAACGATAGTTCAGGATAC 201593
201672 201692
201672 TTATATGCAAAACCAATGAAGGCGACAGAGTTTATTCATATGTATGTAAACAGATAATA 201593
201732 201692
201732 ATTAACCAATATTGGA 201593
201758 201808
201758 ATTGCTTAAATATTAACAAGCCATCAGAAGTTATTAACGTGTGACAAACGCGAGGA 201593
201818 201808
201818 GCACAATGCTGTTAGTGGAAATACATATGAAGCGTGTGCTCTG 201593
202823 202873 202876
202823 CATTGATTTGAAGCCTTTACGGGCGACCGTTCTGATACATCATTAATATGCTGCTG 201593
202883 202901
202883 AGCTGATTTTCAGCAAAAGAACGCTTATCATCCATATATTTGATATAAAGTGTGCTG 201593
202943 202959
202943 CGACATCGGCACTGCTATCCGCTCCAGACTTTTCAAGAATGATGTGTGACAGACAC 201593
203003 203007
203003 TATAAGGTTATCTTAAAGTAACTTAAGCATTTTCCGAGAGATAAGTATACCTT 201593
203300 203320
203300 ATTTCATTAATGCTTTCGACGTGATGTGAATTAAGGGTTTTTC 201593
203532 203582
203532 GTTATATCTATTAATGCTAGAGGTAAATATGCTTATGTATTAATCAAGCGCAATA 201593
203592 203582
203592 GTATATTAGATAAAGATTATATAGATCAGCAACAAATCCAGTCTGTT 201593
203965 204015
203965 TTGACATCAAGAAGCCTGATATATAGATTTTGAAGCTTTAACTCATGAAGCGGTGAAG 201593
204025 204015
204025 TTGAGTTTATGACATATGACATTAACAACCTGATGATCATGCACATTAATTAATAGT 201593
204085 204091
204085 GTCTGAGAGCTTTTCGTCAGTTAGATATTTCTGTCATTAACGCGGCGGTGTGAACATTA 201593
204145 204168
204145 ATGTCCTATGATTTCAATCGAAGACGCGGAGTCGATGATTAATTCACACTGAATG 201593
204205 204168
204205 CTGCATTCGATATAGACCA 201593
```



```
...
195664 TGTGATATGCAATTCACAGTATAGTGTGTCTCTCTGTCCACCGCTCATAGGTCGTTTTC
|-----|
195724 CCAGCTAGCGTTGGCGCCGAGCATCATCAGTCATTCACGTAATGCGTTCGCCGTAGCC
|-----|
195784 GAGTTGGGTGACGACATCCGTCGCATATATGTTTTTTCAGAGGACCAACACCTCGCT
|-----|
195844 GAATA
...
195867 CATAGCTTGCATACAAATGGCTTGCCTCAATGCCGGAACATATCGAACGTTTTA
|-----|
195927 TAGGTTGTAGTATGGGATCGTGTCTGTATTTTGTCTGAATA
...
195997 CAATATGTGTGAAGACCAAGATCGTCTGACCCCTCAAGTGAGACTTCATGACGCTGTT
|-----|
196057 ACTGTGCAAAATCAGTGAAGGCCAGAGACCCGTACTTACGC
...
196141 GTTGCCTGTAAACACCATTTTTTCATTCGCCGTATGGAAGGCGGTGAGAGCGCTCCAG
|-----|
196201 TTGCGAACCAATACGCCAGTTATCATTAATCATACAGCCAGAC
...
196388 TAACCGATCCAGCAAAACCGCGCATTTATCTTCAGCGGTGTGAATTAACGATGTG
|-----|
196448 GTTAAGTCGACATCATGTTTACACTATCCGGGCTTGGGCAATCAATTCGTGTAGCCA
|-----|
196508 GCGATACGAAGCTCTGCAAGATTATGTACATCAACCGACGTTTAAATGTATACAGCGCG
|-----|
196568 GGATCTTGGCGTTCACGTTCCAGCAACATCGTCGTTAAGACAGCTGCTGCTGCCATTCT
|-----|
196628 TGTAACTGTAATGCTGTCCAGGCTTGTCAACCTCCAGATTAAATTTACGTGGCTGAT
|-----|
...
197116 TGCACCCCAATATTTGCCCGGAGAAATATTTTGGCCGTCCTTTTAAATCGCTGATA
|-----|
197176 GTGAGAAATACGCTTTATACGATCCGCGAGTTAATACGCCGCCAAGATATCAACCTG
|-----|
197236 AATACGCTGTACTGG
|-----|

197278 CAGAAATTCMAATGACAGCTATTTGGCGAGGGCGCGATCGCAATGCAATTAACGTCAC
|-----|
197338 TTCACTCCGCGTAGGCATTAACGACAGCTGTACAGGTTGGGCAATGAAATATCAGCGCAAT
|-----|
197398 ATCTGGCGTTAAATTTGCAATCGTCATTCGCGGACGCAAGTTGATTATTACGTAATGCC
|-----|
197458 CAGGTATCTGTGGATATTTGTGCTAGATGCAATTTTCAGGTAAATGACTGTGCAATG
|-----|
197518 CCGTAATTCATCCTGATGACGCCCGCCAGTTTATATAGATTAAGGCTTCTGCGAGTAAT
|-----|
197578 GGCTTTGTCCGGT
...
197652 GGGTAAATATTTGTCCCGCTGTATATCCTTATTTTGGCGTCCAGAGAGCGCTTTT
|-----|
197712 GCCACAGTGTAAAGCAGTTTTCATTTGTCAGGTTACGATTAAGGACGGGACAGCTG
|-----|
197772 CATTAACACGCGCTGTGTAATGCTGATGACGGTAGCGGTTGAAA
|-----|
...
197915 AAAGCTGGCTGAGTATTTACCTTGGAGCTTCAATTAATCAATGATCATATGCGCTTAATA
|-----|
197975 ACAGCACTTTCGCCGCGGACTCGCTGCTAATAAATGCGGAGTAA
...
198708 TATTGTTTGGGCTTCTTGGTTTGGTAATAGTTGGTTAATCAGCCAGCGCTTGAAG
|-----|
198768 AAGTTGTCACGATCAACCTATCTGATGTATATGATTGCTTTATT
...
198913 TTGATTTACTTTGTTAAACCGTCATTAATCATTCAGCAAAATATTAGGAGCGCTTAC
|-----|
198973 AAGCAGTGTGTCCAAAATGACATTCGAATCATTAATTTGTTTGT
...
199094 TTTTCCAAAATAATGACTTTTGTGCAATAGTTTGTTTTGTGGGCGGCGATAGTT
|-----|
199154 GCTCACAATCTTCTAGTCATTATGAGAGCTACAAATTTACACATTTG
|-----|
...
```

191734 CCGCGTGGACGTTTCGAAGAAGATGAATGGTTGGATCGTCGCGTTGCGAAAAAGAGA
191742
191794 TTGCGTACGTTGGCAGCTTGGCAACGTACGCT
...
191855 TCGACAGCCCCGGGCGAACCGGCTTGTGTTTGGTGTTTATGCCCCCTGACTACGCGCTTTT
191905
191915 CTGAACAACCATTTTATTTGCCCTGGCTGGTGAGTGTACGCT
...
192038 ATTGGTAGGCTGCATGATGCTGCTGTTTGTGMAAACGCAGCTTATTGTACAGCGCCAGAA
192088
192098 CAATTAAACGACGAGCATTCGCCACTGCCAGCAAAAAATAGACTG
...
192345 GGGCGCGTTGGCTTTTATAGCATCAACATGACTCGTGAATGATACCAATGTGCGTTGCCA
192395
192405 GGCCTCAGCATCCAGAAAAATAACCGGAAACCAAAATATCCAGAAAGCGATGAACTCAGAT
192449
192465 TATGCTCATACGATTCCTGCATCATCAGGCTGACATTAATTTGCAGTAACATTAACGTAC
192474
192525 ACAATTAATATTCGGGCACTATGTGTCCGACGATATGTGTCAATTCAATTATTTAAGGTA
192553
192585 CACCGGCAAGTTGGACTGCGTAATTAATGAACCGACCGCAGGTGAAGGCCCATATTG
192644
192645 TCGTCAGGCTGTAATTCAAAAACAGCGGCCACATTCGAAAGTTTTCTTTGGGCCA
192649
...
192703 CCTGTGCTATTTTGTGAGGAATCTTCTGCAACCGCCCTGAGCCACGCGCGCTGTTT
192753
192763 CCACAGCCCTTTTAAGGTTTCAGGCAATTAATATCCAGACAGAGTGGCGCTGCTGTAAAA
192776
192823 AATGCTCCACATGATTAACCTGACAGTTCCAGCTAATATCAATATCTTGGTGAATCATATC
192824
192883 GTCACTCGAGTAACCCACTTCTGCCAGGCGCTGGACGAAATGCGGCAATTAACACCGGA
192885
192943 AACGTAATAATAGTTTCCATAGA
...
193088 TACAACATCGGTTCCACAATATATGCGCGCATGCGGTCTAATTAACGCATTCGCCATCA

193148 ATGCACACCAATATTCACTTTTGGCGGCGGACGCTCCGGTTTAA
193138
...
193279 GGTATTATCTGTTGAACCGTCATTTACGGCAATAACTTCAMTGTTCATTAACGCTGTGC
193329
193339 TAAAGCGCGGTATGTTTCTCTCAACGTTTCTCTCATTAATAACAGGAATGATAT
193363
193399 GGAGATACAGGATTAATCTTCAACTGGGGAGCTGTGCTGTTTCTCCCCACGGCCACG
193458
193459 GCGTTCAGGATTAAGACCCAGAAATAGACCGCGCAACAAATCCACATTAATGACATTA
193463
193519 CGGCCAGAAAGAAACGAAC
...
193560 GAAAGTACGCTACGCATRAGGGTATGCTAATACACGACATTAATTAATAAACGCTACGA
193610
193620 TCGATTAATCATTTTTCGGATACAGGCTGTGMAAACTCAGAC
...
194474 ACGTAATTGCCAGGCAACCGACTAAAAATATCTGCTTTCATTTGGTAGCAACGATTTGG
194524
194534 AAACGACGCTCTTTGACACGCCATCACATCGGGATCAGCAAT
...
194715 TCCTGTACGTAATTAATTTGCTGGGCAACGCTTTTAATGAGGATTAATGGCGATTAAT
194765
194775 ACCCGGGAATGAATCCAAATTCGACGCAATTTGCCAAACCTGATTCAGGGGTGAAGAC
194835
194835 ATGTCAATAACGAGTTTTTTTAATTCCTCTATGCGTATGCCATTGCTTCCCATTAAGGC
194865
194895 CAAATAAAAACGTGTGATTACCTC
...
195126 ACTTCTGCACTTGTGCGACGTGGCAAAATATTCTGATGACCACTCATTCGCCAAT
195176
195186 TTTACTTGTATTATCCGCTGGCGTATGACCCAACTGCGAGGGGG
...
195398 GCGCAGCCAGGCAATTTGTCACGCACTGATGTCCGACATGAATTAATAACGCTGTC
195448
195458 GGCAGCTTTCGTTCAACGTTATGCCATGAATGCTACAAAAACCATTAATGCGGCCACGG
195473
195518 TTGCTCGGCGAGTAAGATTTGCGATC

```
188881 CCACACCGATCTGTGGGATTTCCAGGCTTACGTTGAAGGCTGCAGAAAAATTGTGATCT |-----|
188941 GCTGCGTCCACAACAGCAAAAAGC 188909
...
189156 AAGCATTTAGCAGTATGAAGATAAAAACGCGTGAATGAACCGTCACGCCGAGCTTACT |-----|
189216 GAAAGGCATTTGTCGCGCTGGCGCTGGGGAAGTTGTCCGGTCGTCATGCACAAAAAAC 189234 |-----|
189276 GCAAGTGGCCGGGTACGCTTTCACCGGTTGCAGGCATGAAAAACACCGTTTTATG 189256 |-----|
189336 TGACGATCAGCGAGGATCTTACGCCACACAGGCCCAATGATGCTGTGCGTTGA 189357 |-----|
189396 TCTGCTTGCAGCGATPAAAGCGGATCTTGAACGGTTTTCGCTTGTGACTCAGCTT |-----|
189406 189406 189449
189456 TGCCTTTTCTGACTCAGGGCGGAGCGCACACGAAAGCCCAATCCGCGCTCCACACT |-----|
189489 189489
189516 CGATTCCGGCATCTTGGCGGCTACATTG
...
189546 GCCCATATCTCACCATCAGCTTATCGGTGGGTCACTCATGTTTGTGATGACCGCTTG 189536 |-----|
189606 CCTTGGCCACAGATGCCGAAAAAGTTGCAGAATGACGCGTTTCCACAGCAGCT 189655 |-----|
189666 GGATCGGCGTTATGTGATGATGTGTTGCTACAAATTTGCGCCACACCCAGAGACAC |-----|
189672 189672
189726 GG
...
189787 AAGCGGAAGGTTTATTTCCGATCAGCGGCGGTAGTAAGGCAAGAGAGAGCGGATTT |-----|
189847 AATTGCTGGTTTCAAGACGCGCACTGCCAATCCCGATAGCCAGA 189837 |-----|
...
189964 AGCTATCAGGCACTACGCTTGATTCAGTTTCGATGGAATTTTGGGACAGACGCCGCTG |-----|
190014 190014
190024 AAAGAACAGCAGACGATTTTGGCGGTGATTAAGCAAAACGCGTGGCGCGGTGAATGCGAG 190035 |-----|
190084 CATGAGCATGATGCTGTGATTACGCCAGCGACCGGAGAGGAGGAGTTCGCTGAC |-----|
190111 190111
190144 AGCCATATCCGGCTGGCGAATCC
```

```
190175 CGGAGAGTAGTTCACACCTGATGCTCGTCGTGGCTACAGTTATTTCAGTGGGCTACCA |-----|
190225 190225
190235 ACTCCGGCACTGCATATGGGGCTACTGTTTGTCTGTTACCAACAGATCTGAAAAAG
190295 GCTTCCTGACAGTACAAAAAAGGCTCATGCGCAAGCCGCTGGAGGAAATACGTTAAACCTA |-----|
190329 190329
190355 TCGCGGCGGTTATTTTTTTTGGCGTGCAGGAGTGAAGGACCGCAAGATTTATCTCGAA |-----|
190399 190399
190415 GCGCGTTATTGGGGGTTTATGTTTATAGACGGATPAGGCTTCTTGGCAGATGCTGA |-----|
190452 190452
190475 TCGATGCTTGGCGGCTTCTTATGCTPACAAATCGGTGCGGTTGCGTAGCGGATPAAAG |-----|
190533 190533
190535 CGTTCACGCGCATACGGGATGTCGACTGATGCTGTGCAAAATCTGTGAAGCAC |-----|
190538 190538
190595 ACAGCTTTTTC
...
190733 GCGAAGTTCCCGGAGTAGTATGTCAGTCACTATGAGATCGCGGATGATGATGCTGT 190783 |-----|
190793 ACTGACATGTTTATGATATATCAACGCGCCACACTCGCGGAGTTT
...
190854 ATCTGCTCACTTTGCAATTTTCAACCTCATCTTTCTTTTATGTTTACCGACCGCTTA 190904 |-----|
190914 ACGGTGCGCGTAGCGTTTTTTCGGCATTTTATCCAAAAAGCAACTA
...
191325 GCGCAGCAAAAGCGCGAGAGGCCCTGATACATPAGGATGTCGACAGATTTATCTCACCC |-----|
191375 191375
191385 GTTCAGTTCTGCAAGCGGATGAAGATCTTGCTTCTTACTGTGAGA
...
191434 CGCAGAAAGTTTGTCCCTATTTTTCGCCGGTCTACGACGTGCTGCTCGGCGCTTAGG |-----|
191484 191484
191494 GCGTTCCTTTATGACAGTACTGCTCGACCGGAAAAATAGGGAAGTGAATTTGCCCGTT
191554 CGCTATATGCGTGAAGTACTCTTTGAAAAATGCAAGTCATTTTGAACGAGCGCAGAA |-----|
191588 191588
191614 TGTGACTGCGCGCAAAATGAATGTTTTTAAACCGCGCTCGGAGAGAACTGACGCTTAT |-----|
191672 191672
191674 CGTCAACGCTGATATACGCAATGCGATTGCTCGCGCGCTGTGCTCGGATTTAAGTGA |-----|
191677 191677
```

185978 CTTACAGGGTTAATGATTCGTATTAAGTCGTAGCCATCCGATGGCTCCGGCGATTCAG |-----|
186028
186038 TCGGGCGACACGAGATGTACACTTCGGATGCCCTTCTTGATGTAGTAA
...
186244 GTTTCATAAAGTATATAGTTACATTACCTCGTAATTTTATGCAATACCTCGACGTTAGCC |-----|
186294
186304 TTCAGCCAGAGAAAAATTAACTCGCAAAATTTTAAATTATTCGTTAGGGGAATGGGTAA |-----|
186343
186364 TTATATTCGACCCCTAATTTACAGGGTTCATCTGAAGCAGTGGCGTCTCATCTTATTAC |-----|
186406
186424 TTCTCTTCTGTATTCATGCTAGTGGGTTTGTGATAAT
...
186681 ATATATAAGGGGAATTAATTTTGGGGATATTAGGCACTATCAAAAATCGCTCATTT |-----|
186731
186741 CCAGACTTAATACAGCTTTCTCTCGTTTAGGTTGGGGCAAAAA |-----|
...
186804 GGTATATTTCAATTCACAACCTTATCACTATCTATTATAAGCAAAAAACACGTTCTTT |-----|
186854
186864 TGTGAAGGACTAAGTGAATAATTCACCCTTCGGGACGCGTCACAGGAGTTTGTGGCGAC |-----|
186916
186924 CAACATTCGTACAAAGCTTCACTCTGCGGGCGAAGCGTCAGCTAAATACCTTAAATAATTC |-----|
186956
186984 GCTTTTCTCCGGATTGTGATCTCCTCATTTTACATTCACAACATTTTATATAATATTT |-----|
186986
187044 CCGCTGTATTTTCTTCTTAGGTAATGATTAATCGCTATCACTGCGATTACTTTCTTTG |-----|
187074
187104 CATAGATTGACTCAGAAAAACGTTTAAAGGNGGTGCGCATGTTGTCCGTTTCTCATTA |-----|
187122
187164 TGTTCGCGGAAGA
...
187323 GCGAATTTCCGCAAAAAGACGAACTGTTTGAAGTATCGTGGCGGATGATCGGCGGTG |-----|
187373
187383 TGATCCTTACCTGATGAGTTTCTTGATGCGCAAGGTGCGCAAGCTCAAGATACAAC |-----|
187427
187443 TGGAGCAGGCTGTGATAGCCGATTGACGCTGGAATTCATCANTGGCTGGGCGTGGTGA |-----|
187492

187503 TGATGGCTTTTGGCGTTGCAAGGGAAGGCTGGATGCGCTCT
...
187662 GCCTCAATCTTGTGCAATTTTAAATGACCAACCTGTTTATTTCTTCTGTCGCGCGAG |-----|
187712
187722 GGCTGGCAGCTGTGCAATTCGCCATTTTCATGAAGCCGGATTTGGAACCACTTTCAGG |-----|
187782 AATGCGCTTTCATATATGAGTGGGTCCTCTCACTCACTCGCTGTTTGGCAGCGTATGG |-----|
187784
187842 AAGGATTTTGTGCTATCAGAAAGCGGAGGTGTCAGCGAAGTCGCGCTGTGTTATTT |-----|
187864 187871 187883
187902 ATCTCATCCCGCGGTGGCATTTTGTCTGCAACGACGCGGAGGCGGACGCGTCTC |-----|
187912 187948 187954
187962 GCTCCGCGTAACAAATACGACGCAAACTCTGTTAGTTACAACATAC |-----|
...
188052 ACGCATTCGAGTTGAGCGGTGCGCTGCTGTTTCTTCTGCTTTTATGCTAACGCGCTG |-----|
188102
188112 ATGTCCCGCAGGTCAAAGTGACCGGTGACGATTAAGCAGTGGCAGACCGATGACCATTAAGG |-----|
188172 TTAACGCGCGAAGAACACAGATTCAATTATTAGAACCCAGACGAGAGGCTGTGAATGG |-----|
188174 188218
188232 AGATCCTCAAAAGCGGTGATGTTGTGTGAAGAGCGGGAATATTCGCCCTTGCTTAGCC |-----|
188273
188292 AGAAATGACGCGGAATTTACAGCCTGCGCAATACGATATGACCTGCGGTCTGTGACTA |-----|
188352 ACCCGAAAGGCAAGTTGATTCCTCAAGGTGAGGCAACGCGGATGCGGCGCAAGTGATG |-----|
188369
188412 CGCTGTAAATCT
...
188469 CCACACAGCTGTGACCGACACCAAGCCTTTACGACGCGATTTAAAGCAGCGCATATCG |-----|
188519
188529 AAAAAGCGAAAGCACTGATATGCGCGGCGGCGGACACACTATGAGGCAATGACGATTG |-----|
188589 CTGAAGTGTCTCCGATCTGATGACGACATTTGACGCCCGTGAGATGATTACGACAA |-----|
188621
188649 AAGCCGCTGATCCAAATTCACCGGTTTCCACCGCTCTGAAAAACCATTTGTTGGCGCA |-----|
188702
188709 ACACCACCAAAAGCATGATGATGATGACGCTGACCAAGCTTTATTAACGATGT
...
188821 AGCCGACTGATTGAGAAAGTGGACGACCAAGAAATCAGCGGTGAAGAAGATCCGCTACAG |-----|
188871

```
182737 AAGTGCAACCGCAAAAAATGTGAGAGAGTGCACACTGATGAATAATAGTGTGCGGTGAC |-----|
182787
183797 ACTAAATTTAATGTAATAGTGTGTTAAATGATTTGATTAACCAAGCGCTTCCGGCAG |-----|
182843
182857 GATACGGTCGCCCTGTGTAAMACATACTGTGTACCCCGTTCGCGTGGCAGATATTAACG |-----|
182863
182917 GC |-----|
...
182979 TGTGTGATCATTTTGTGTATATCTTTGGCATGATATTTGATTTGGTTTATCGGCTGCG |-----|
183029 183036
183039 GATCAGCAAAAACTTTTGACGACATATTTCTGGGGGAGCGTAGCCTTGGGCCATTGCTGA |-----|
183041 183056
183099 CGGCATTATCGGC
...
183141 TAATGGGGTTCGCCGGCGCGGTTTTTCTTCCGGGATTTCCGAAGTTGGATCGGCATTG |-----|
183191
183201 GCCTGACATTTAGCGCGCTGTGATTAACCTGGAAGTGTGTGGCCGGCGTTGCGTGTGCAVA |-----|
183261 CCGAATACACAAATTAACGCTTAAACATTACCAGATTATTTCAACCGCGGTTTGAAGATA |-----|
183275 183306
183321 AAAGCCGATTTTGGCGCATTTATCTCCGGCGTGTGATTTTG
...
183416 GAAGTACCTTTGGCATGAGCTACGAACGGCTCTGTGGCGGAGCTGCGCGCAGATC |-----|
183466
183476 CTTTACACCTTTATTGGCGGTTTCCTCGCGGTGAAGCTGAGCTGACACTGTACAGGCCAGC |-----|
183469
183536 CTGATGATTTTTTGGCCCTGATCTCTACGCGGTTTATCTGTATTAATCAGTGTGCGCTTT |-----|
183559 183569
183596 GGTGACTCGCTGGAAGTATCAACAACAAAGACATCAAAACGTGATATGCTCAAAAGG |-----|
183656 CTGAACCTTTTGTCCCATTTATCTCGCTGATGGGGTGGGGGCTGAGTTACTTGGGCAACG |-----|
183665
183716 CACATCTCGGGCGCTTTTATGGGGGGGATTTCTACCCACAGCATCGTCCAGCGCTGCT |-----|
183758
183776 ATCAGATGACCTGATGATCTCTGCTGCGACAGGGCGGTGCTGTGCGCTTCTTTGAT |-----|
183836 ATGCTTACTTTAAGACAGACCGCTGGGTAGCTGTGCGGTAAACCAAGCGCCGACGCC |-----|
183836 183884
183896 GTGTTTATCGAAGTGGCGCAAAATCTGTTTAAACCGTGAATGCCGGGATTTCTGCTGCG |-----|
```

```
183955
183956 GCGATTCTGGCGCGGCTAATGTCAACGTTAAGTTGCCACGCTGCTGTATGTTCCAGTGGC |-----|
183960 183979
184016 ATTACGAAGATTTGTACAAAGCGTTTCTGCGTAACATGACCGCAGCAAAAGAACTGTG |-----|
184036
184076 TGGGTAGGCGCTGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG |-----|
184105 184114
184136 CCGAATAACCGCGTGTGCTTATGAGCTACGCGGTGGGCGAGGCTTTGGCCGCGGCTTT |-----|
184189
184196 GGTCCGGTGTGCTGCTTTTCCGATGTGTGTCACGATGACGCGACGCTGCGCTGCGC |-----|
184256 GGGATGATCATGCGTGCCTGACGGGTGATCGTCTGGAACAGATTGCGCTGGGCACTG |-----|
184283
184316 TACGAATTTATTCGGGCTTTAT
...
184343 GGCACATTTGGGATTTGTAGTGTATTGTTGCTGGGTAAAGACCCCTGACGGCGATGCA |-----|
184393
184403 AAACGCTTTGCCAGAGCGCATGCGCATATCATTCGCTCCGCGCTCAAGGTTGACG |-----|
184404
...
184855 ACCCATTTCTGCGCCCTGTGAGCTTCAACTACCCAGGCGGTAAAGCCCGCTTCTTC |-----|
184905
184915 AGTAAATTCCTGGCAGTAGTTGGCCGAGGTTTCTTCTTACAC
...
185116 TTTTGAATTAATAATGATTCGCGGTACACCATCTCTGAACCTCCCTGTAGCGCTTCCG |-----|
185166
185176 GCTGTGTGTCATAGCCGCCATTTTGCATGTGCTGTACAGTCAATTAACATATTGC |-----|
185208
185236 TGTGCTCTGCTCCCTGCTGCTGTGCTGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTG |-----|
185289
185296 TTTGACGATTTTGGTGGATGTGCTGTATTAACACGATTCTGTAAATTCACATTCGTTA |-----|
185349
185356 CGCCAGCGCAGCGAGACATATTTCCAAATTTCTGCAATGCAATTTGTTGG
...
185822 CCGTGTGTAGTCCGCATCCCTTTGATATAATTAATTCATCTGACACGAGACGCTCAAG |-----|
185872
185882 TTTCAGACAACATCTCGGGAGTTGTTGGGTTGATATATAGTTGC
...

```

180100 TCGCGGCAATCAGTGGCTTTGTCTGCGGTTTTGGCCAGTACGCTGTACCTGCGCC
180115
180160 AGTGGGAGGATCTGCCCGGTGAATAATCGCCAGCGGGAAGTTCCAGGGCTGATACAT
180168
180220 ACCACAGGCCCTAATGAGCATGGGTTTCGTTAGCGAATCATCCCGCACCTGTGCGGCG
180236
180280 TAGTAGTGGAGAAAATCGACCGCTTTCACGCACTTCGGCAATGGCGTTACTGAAGTTTTT
180321
180340 CCGGCTTCAGCGACCAGAAATACAAATCAGTGTCTGCATCTGGCTTCCATCAGCAGCGCA
180400 GCGGAGTGCAAAATCGCCGCGCTTCAGCCGAGAGTGGCAAMCCAGATTGGCGCTTA
180412
180460 TTACCGCACTTTCAGCGCCTGCTTACTTACGCGGCGTTCACGTACATAGCCC
180475
180520 ACAATATCTTTCGGTTCGCGGGTTAATAAGGGGCGACATCTCACCTGCCGTACCGGT
180553
180580 TGTTCACGANTTGGCAAGGCTGCCATT
...
180742 CCAGTTTGCCCTTCTCTGTTGTGCCAGTTTCTTACAGCAGTACCGGATCGCGCACGT
180792
180802 TCATCCAGTGGCAAGAAGGTATCGGCATACGTTAACAAAGAGGTGTAGCACCGTTT
180862 TCCAGCAGCGCGCCACCAGATACGCCAGCGTTTCATGTGTGCCGACCGGGGCAATA
180892
180922 ATACGACACGAGCGTTAAGTTTGGC
...
181004 ACTGGAATCTGCTAGACCCGAAATAGTTCGTGACCTGCCAGTTGATTAATCGCCGCA
181054
181064 GCGTGTGGCGTGTGTGCGTTCGGAATGCGGGTAGATTAGATTGCGCACCGCCAGCAGCT
181071
181124 TTTTCGACAGCGCAGATTAAGAAACGTGCTATATACCTTCGGGATATAACCGGATAC
181146
181184 CTTCAAGCGCGTCATCT
...
181209 CTTGATTTTCACTATCCAGTACGGCCTTTCACAGCGGAATCATTAAGACGCGCT
181259
181269 GCGGTGGCGGATCAATCAGGTATCATTCACAGCGGGGAGCGTTTGTATTAAGCCTG
181310
181329 AATTACAACAAACGATGCGGTTCAGCCTGCCAGCTGTGGCTGAAACAGAGTTTTCAG

181389 CAGATCGAGGAGATCTCCACAGATTCGCGCTCTTCGGCTCAATGTTGATACCAATATC
181409
181449 GTACTGACCGCCACAGCAGGTGAGTGAATTCA
...
181490 AAAGCTTCCATTACCCGGTCACTACTGGGGCGGCTGTAAACGGGATGCACAGCGCACA
181540
181550 GCTTGATTGAATGCCCGACCTTCATGATGCCACAGCCGTTAGACGCTTTACCAATGG
181594
181610 CGTGAATGCCCTGTGATAGGAAMCANTGACGCTGTGCGTTCGCGCAGTCAGTCAG
181615
181670 C
...
181682 TATCGTAGAGTAACGAAACCTTCTCTTCCAGCTTCGGGCAATGGCTTAACGCTTCG
181742 CGATGTTTCCGCCGTGACGACTGCTGCCCATCAGCGCATTTGC
...
181973 TTTTGTGCGGAATTAACGATCAGCGGTGCTTGTGCGGAATACGACACAGCTTCTG
182033
182033 CCAGACACATTCAGCGCCACGCTTCCTGCATGACAGGAAACCTCTGCAATAACCCCT
182044
182093 GGACCAT
...
182187 CGGCAAGCGGCTTGTTCACAGCATAAACCCTTCGTTCCGGGCGGCGATAGGC
182247
182293 GCAAACTCGAGGAATGCGTGTGTGTTCTCTGCGGAGTGGTCTTCACTGCTCTCA
182353
182353 TTGSCAGCGCCAGAAAGCAGCGAGTAGCTGTGAGAGATGCTGTTTTCAGTTGT
182358
182413 TCGAGATTAGAAAAAATCGCCTGCTTAATTAACCAGTGTGCGTGCATGATACGTGTC
182428
182473 GCGCAGACTTAATATAGCCTCAGCGCTGCGTGTGTCAGCTTAACCCCATGCTGTGGTT
182502
182533 CCGATGCCATTACTCTGTTGTTCTGAAGGTGCACTTAACGTTATCGTAATATCA
182593
182593 TGATGTTGCAACTTGTGCAACCATGTTAATGTG
...

```
177878 CTACAAGAGGGAGAGCCGATACGCAAGGCGCAGTGAACAAACACGGGTAAACGTTCCGCG |-----|
177938 CGCAGTAAAGCCGGAAGAAAAAGCGATTCCTTAGCCGACACTGACACTTTTTCACAAATT |-----|
177998 CGGTTTTCACGGCACAC |-----|
...
178035 TTGGCGGGTGTTCAAAAACCAATCTGCTGTATTACTTTCCGTCAAAAGAGCGCTGTAT |-----|
178085 |-----|
178095 ATTGCCGTGCTGCGGACAGATTCTGATATCTGGCTGGCACCGTTAAAGCGTTTCGTGA |-----|
178142 |-----|
178155 GATTTCGCCCGCTGGCGGCGATCAAGAGTACATCCGCTGGAAGCTGGAAGTCTCAGCG |-----|
178172 |-----|
178215 GATTATCCGACAG |-----|
...
178240 TCTGTATGAGAGTCTGCGACAGCGCGCGCTGTATTATGATGAACACTGAACGCGGATTTGA |-----|
178290 |-----|
178300 AGTCATTAAATTGATGAGAAATCGGCGCTGATTGCCGTTGGGTCAAAAGCGGCACAACTCG |-----|
178322 |-----|
178360 CGCGGATTGATCCGACAGC |-----|
...
178404 ACTCAACATTACGCCGATTTGCCCCCTCAGGTGAGGCGGTGACAGGCGCAACGTTGCGC |-----|
178454 |-----|
178464 GATGAGTATTTTCAATCAAAACGGTTGAAAAAGTGCAGCGGATTA |-----|
...
178532 CTAAATGATCCGGAAGAGGGTGTAAACATCTCCGGTACTGTATTACCGATCGTCAATTA |-----|
178582 |-----|
178592 AGCTGGCTTACCCGACGCGGAGCGGATTTCACACTAGCGAAGCGCTGATATACAGCC |-----|
178596 |-----|
178652 GTTCCAGCAGGATATTGCTTTCGCCACGCGGCAAAACCC |-----|
178634 |-----|
...
178809 CGCCAGTTGAATACGCTGCTGACAGTGAAGTGAAGCAATGCCCTTCACTAACTGACGATGAG |-----|
178859 |-----|
178869 CGCGTATCTGGCCACAGACCTGGCTGCCACCGCCACGACGCGG |-----|
...
178928 GCGCATCTGTTCATCATCCGCGATACACAATACGCGCTCAAGCGGACGACGCTCCAGG |-----|
178978 |-----|

179888 TGTTCGCTACACCCGCTGCGCCCTGGACAGCAATCGTTGTGTCTCCCTGCGCCAGCTCGC |-----|
179018 |-----|
179048 CATATTGGGTACATACCGCCCTGCAATTCCTGGAGCATTTGCTCCCATTCGCCAGTGCAT |-----|
179062 |-----|
179108 TTAGCGGCTGAATCAATGCGGCTTTC |-----|
...
179148 CGGATACCTCTGCATCCTGACGCGCGAGTGTACCCGCCAGCGCATTTCCGGGCGAATTCGC |-----|
179198 |-----|
179208 CAGCAGCGGTAGAGATAGAGCGGACCCGCTCTTCCGCCCGGTA |-----|
...
179305 ATATTAGGGTTACATACAGTTTACCACATGCGGTGAGCCAGTACCTGGGCGGATGGTT |-----|
179355 |-----|
179365 TCATCAATACGCGGTATGAAACGCCAGCGCTGACACCATTAGCCGGAAGCGTTAATCTGCTCG |-----|
179382 |-----|
179425 ATCAGCTCCGGCAGCTGATTACGTTATAGCACCACCAT |-----|
...
179471 CCGGACCAAGAAGACCTCTTTTTCGAATTCGCAAGATCATTCAGTTGATGATACAGCGTGGGG |-----|
179521 |-----|
179531 CGACAAGGTGCGGCTTGGCATTCACGCGGCAATCTTGCCTGTTTCCGACACCGCCTGGA |-----|
179534 |-----|
179591 ACACCGGAGCGCCTTGTCTACGCAATGCTGTGAATATGGGCTCAATATTGGCTTTCCTT |-----|
179627 |-----|
179651 CGCTATTCATACACCGGACGATATCGGTGTCAGGCGAACCAGATTACCCATCCGGCATTT |-----|
179684 |-----|
179711 CGGCCATTGCGCGCGGACAGCATTTTCAACGTGTGCTCGGCAATCTCATCTTCGACGACACA |-----|
179771 |-----|
179771 GCACGCGGAGCGCGCGACAGACGCTGACCCGCACTGTGCAAGCGCGGAGGCGACACATGTA |-----|
179779 |-----|
179831 TTACGACTTGTGCGTCAAGTGTGAAGATGACAAATCATATGCGTTCAATGCCCGCGGTTT |-----|
179891 |-----|
179892 CAGCGATGAGCGGAATFAGGCGGACCGCTGACGAGGCGGCTGGGATATTGCGCTGCA |-----|
179909 |-----|
179919 |-----|
179951 GTACGTAAGCGACTTGTGTTGAACCGGTAAACATGACCCCGGTACGGGATCATCACC |-----|
179958 |-----|
180011 TCA |-----|
...
180040 CCTGCTGCAATTCACACAGCGCTGGCGGTACGCCGCTTCAGAGCAAAATATGCGAATCCCT |-----|
180090 |-----|
```

```
175160 ATGGCGGACGACTGGCTAATGCTCAATCCCTCGGGCTAACGACGGCTGATGCAAGTTCCGC |-----|
175220 TGCCATCTGGGTGATACTGTAACTCTTCTGCCAGCGT ...
175384 CATCAGCATTAAGGGGAGGTGAGAGTGAAGTTTCATCGCGGCTTCTACTTGGCGCATATGC |-----|
175444 GCAATTGTGGCATTTCCACGACGCGCTCAGGTTTTTACCAGTCCGCACTGAATGCAAGAT |-----|
175504 CGCG 175452
...
175763 AGGGGGCCAGCGGTGCGTGGCTGCTGCCAGGGGAAATAATTACGATTTTGGCATCGTTAAGC |-----|
175823 TCCTTAAGCGATACGAGCAAGAGAGCTGGAGAAAGCGCGTTCGCAAGCTTTGACGCTC |-----|
175829 175829 175869 175877 175880
--|
175883 GCTGACCCGAGCCAAAAAGTTTCGATATTGAACAACGCGGCTTCTGTACAA
175885
...
175957 TTGGCTCTTCCAGCACGACGCCGAATACTCCAGAAAAAGCATCGGCTAGCGTCAATT |-----|
176017 CGACGACAGCGTTGGTAGCGATGCTGTGAAAAACGAGATGGGATATTCGCGGGCTGCCGA |-----|
176024 176033
176077 AAATGCTGTCCAGCGGGCTATTGAAGAAACCGCTGTAGCGCGCTTTCGGCAGCACAAATAT |-----|
176137 ---| CGCGAGGCTGGGGGACACAGTTCAATCCACAGTTGATTAATCCAGAGGCTTTCGCAGCA
176140
176197 ATTTCCTCGAGCTGCGGCTGTTTACGCATGTTTTGAGGCGCTTCATTTATGAATAT |-----|
176257 TCGGTGAGCCAGGGCCGACGCTTCGACATACTGTT 176237
...
176385 GTTAGACATCAAAACCGGCGAGATCTAAGTAGCCGCTGGTGTGGCATAGCGTTTTCG |-----|
176435 176435
176445 AATATCAACGACGATCAGCGCACTTTGTGCGGATCGAAGTAAATGGCTTCGGTGCAGCG
176505 GTTAATGTCTCATCATGCGCACCTCTCTGATGACGACAGGAGATGGCGGCGCACTGCA |-----|
176511 176511

176565 TCAGTGTGTAATGCGCTCGCCGAAGTTTCGATTCGCCAGAAATAATCGTGAAGGTTA |-----|
176625 ACAGCACGCTTGGCACCAGGACGCTTGGCATTTCAATCTAATATGCGCGGCACTGG
176685 CGTAAGAACCGACATAAGCTCCCATATTGATGTTTACCGCGAAGTGGAGTGGCCATCT |-----|
176745 GACGAGCTTGGTGTCACTACCGAGCGGGATATCTTCTGACTTTGTTGCTTAAGCCAGC |-----|
176805 TTAAGCGCTTCCTTCAATCCGCGCCGCTTGTAGTGTTCATTTGGCGGAGAGCGCATGCT |-----|
176807 176807 176860
176865 CGGTTTCATCGGCATACCATTAACAATAACATAAGACCAACGTCGCGCGCGGCTTTCCT |-----|
176865 176905
176925 CTGGGCGCTGTTTCATCCGCGGAGCGGTGCGGCGCA
...
176973 TATTACGCTTTTGGCCGAAGAGTGAATTCGCAATGCGCATCTAGACGAGAGACCATGCGC |-----|
177033 177033 177037
177093 GACATCATTCATGCTGAAGAAAAATCGCTTTAAATTCGCTTTTCCGTCGCCACAGGT |-----|
177153 CGCGCAGCACTGAACATACCTCGGTGAGATGTCGTAACGACGGGAGAAATAGTCATGCG |-----|
177213 CAGGCGAATACCATCTGCTCATCTCGGCTTTTGGCAGCAGTCACGAGTTGAGCG |-----|
177213 177213 177268
177273 CAAGACGCCGCGCAGAGATGAGTCGATGTTGGCGCCATFAGGGCGACATTTGGCGGAG |-----|
177273 177281 177316 177319
177333 GTAACGTTAAGTGGCGCGGCTGGCTAATCTGAATCGCGAGGTCAAGCGCGCCAGCC |-----|
177335 177335
177393 CGGCCATCAAGGTGAAGCATCAAGGTTGTGATCCAGAACTCAAGTTTTCGCCCAAGC |-----|
177453 CACGCACTTT 177407
...
177600 CAATTTTCATCATCAACTCTCTTCTGTCGTAAAGTGAAGTCAAGACGGGCGCTGAT |-----|
177660 TCTGCATATTCCTTTTTCAGCGCGGATATGGCTTTGTTGCAAAAGCGGA |-----|
...

```



```
172797 CAGGCGCATGGCTGACGAGAGTAAATGACGGCTTTTCCAGCCCTAAACGTTGTC 172847
172857 ATCAGCCCATACAGACGACGCGACAAATTAACTCTACAAGAATCATTAACCGCTGATC 172877
172917 ATGCCACGGGTAACA
...
172966 CCGCCATCCAGCTATCAAAAGCCTGAAGCCGATACACTTTTACCGGATAGGCGGTAAAT 173016
173026 TCAGGCCAATCGGCATCAACCAACCGGACCCGTCACACCGCGGCATTAACCGCTTAATCC 173034
173086 ACGG
...
173091 GTACCGATTTTCATCACCAACGAGGCCAATGACGGTATAAAGCCCAACAGGCGCATATC 173141
173151 CCGCCAGCGCAATGCTGATATTCGGGTTAATGCCCTGACCGTTAAAGCCCGCTGCGGCG 173207
173211 A-TACACAGCGCCGACAAAAGCCCGCTGAGCCGAGATTACTGGGCAAGCGCC 173212
...
173317 GATTGGGATCCAGCCCATCAGAAATCGGCATTAAACCGCTCGGCCAACAATCGCCACCG 173367
173377 CGTGTGAACGCGCCATTAATGCGCTGTGAGCAAAAGGCAATGCTCATCCGCGGACCA 173384
173437 CGCGGCTCTGTGAGAGTCAATTTTAACTGCGCAG 173416
...
173533 CAGCGCTGATTAATGCGATCGAACCACACAGCCCGTAGGGGTGCGCTGATGAAT 173583 173591
173593 CGCTTCGATGGCGCAAAAACAAATGTCGTGGGTGCGGACGCTCACACCTGCTGATACG 173596
173653 GCAGTCAACGAAACCAACGACGCTTCCAGTTGCGGCAATCGGTCACACCATCTCGCA 173669
173713 GCGGCGGCGGCAAAAGCGGTGTTCCATGGCGTTTGGCCGCCAAAAGGTTGAAGCGG 173741
173773 CTCCGCCCCGGGCTAAGTGTATTACACACAGCGTTCGATTTTCATTGAATACGGGCA 173782
173833 GACGAGCGCCCGACGATTACAGCACACCAAGTAATGTGGAGAGCGTGTGTCACGCTCA 173837
173893 GACGCGCTGCGGGTGAACCCGCGCGCCGCGCTGACCGCTCGGTGATGATATTGA 173896
...
174033 CAGGCTTCTTCAAAAGACAGACGTGGCAGCGCGCAAAAAGTTTGTCTGTATCGCTAAG 174083
174093 CCGATATTATACAGCAGATTGCTTTCACGCTGCTGCTGCAAAAAAGCGTCTACG 174140
174153 TGTGACGATCAAAAACCCGACATCGGGCGGATATCCAGCCCAAGCGCAGCGGCGATG 174157
174213 ATCAGATAAGCCCGCTGCATGGAACGTGTGCAAAAGCTGTTCTTCGCAAGTTGTGG 174246
174273 CTGAGGATAAACCAACTGCGGCGCATCGCCGTGAGGAAMACGTTGGGTAACCGTTCAATA 174296
174333 AATTCACTGTCCCATGCGAGCATAGCGGTGACGGGCGCGGTGAGGGTTTTCAGATTG 174350
174393 CCGCTGAAAGCGCGCGGCGCAGACGTTCTTTCCTCTGCAGTCGATTAACACGATA 174402
174453 CGTGGCGGAGAACACTTAAGTGAATGTCGCGCCCATTTTCATCAGGCAATAAGCTCCGT 174415
174513 AACGTCATCGCTGACGGGTGTCGCCGCCAGCGTTGTGAGTCGGGCAATCGGTGAAAC 174513 174521
174573 ACGG
...
174631 CTTACGGTGATGTAAACAGGCTGGCAAGCCGTTGAGTAACAGACATTAAAGCTTGG 174681
174691 GATCGTCACGTTGCAAGCGGTGTCGCCCATAGCGCATCACCATTTT
...
174819 ACCGGGACGGGANTGCATCCGATGTGACTAAAGTCAGGCGGTTGAGGGCGTTAAGT 174869
174879 CGACGCAGTAAATTATTTTGGCCTGAAAATGCGGCAGTGCACGCG
...
174980 GAACAAAGGCTGCGCTTCCACCCATGCTGCGCGCCACCGGTATACAGTAATCGTCCG 175030
175040 AACTGAAACAGCGGCGGCTATGGCGGTTTATTCGTAGCCAGCGGTAACGACAGACAG 175064
175100 CACAGTTACGACGGGGGATGATCAGGCGGAGTGCATTCCACACAGCGACGAGGCG 175125
```

```
16906 ACGACGACAGGAATATATCCATTCATTAGCGTCAGGATGCCCGGAGTATAGTCG 169062
      169037
16906 CCCACTTCATTGCCAGTCACCTTCTCTGTCAAACTCCAGCGCAACAAGAAC
      169067
...
169375 TTACAGTTCGGCTTCCAGTCTGCTATGCCGCTCTCTGATGCTTCAAGTAAACGCTGCT 169425
      169425
169435 TATCGCGTAGCGCTTCTTCCAGTTCAGCAACATGCGATTCACATCAATAAGTTGTCT 169443
      169443
169495 CTGC
...
169511 AACTCTAGCTTCCCAACCGTAAGCGCAATCTCTGTTCTCTGTCGCGCGCTTGAT 169561
      169561
169571 GTATTGCTGTTTCTTTCCGTTTCATCCAGAGTTCACAGACGAGCGGCTAGCCTC
      169643
169631 TGTATGAATTCAGCGTTCATAGCCTGAGCATCTGTTTCACACGAGCAGTTAACATGA 169643
      169643
169691 CATTCCTGC
...
169922 TTCATGCGCGTCAATACCGAAGCTTCGAAAAATGCAATCTATTCGCTTTGCGCTGCTC 169972
      169972
169982 CTGCTTAGCGGCTTATTCATTTTTCATGACACAGTACAGCCATCCGCTGCA 170029
      169990
170042 GAACATGATGTAGAAATTAACCAAGGAGCGGTTCAGGCGCTGTATCGATCGT 170067
      170067
170102 TATCGTGTGTGATGTTCACTTCAACACCTCTGTGTTGTGATAAGTTCAATATCTGC 170102
      170102
170162 TGGCACACTGGCGCAAGCTCTGCAATCCAGCAAGCAGCGCTCGCTCATCTATCGGA 170199
      170199
170222 TCGCACACGCGCAACATGAGTAGCAGATGGGCATTACTATCGGATTTGTATTTTGC 170226
      170226
170282 AGGAGAGATTTGCGCTGCATTTCTTCATTTACGCGCGGTGCGATGATATCTGCCAT 170317
      170317
170342 TTCCTTTCTTCAGGCAATGAAAAAGAGCC
...
170714 ACTGATGAAGTATCTGCGTATCTGTGCGCTTTTTCGTACGCTTACCATTCGCGCTTT 170764
      170764

170774 CTTTGGCTGTATCGAGGGTAAGCTGTAGTCTTTTTCCTCGCCC
...
170862 ATTTTCAACGAATCCGTACTCTTCAATACGTTTCTGTAAATCCAGATGCAAGCAAGCTTTAC 170912
      170912
170922 TACTCCAGAAAAGTATGTAATCAACGGGCATTAAACGAGAGAGTGTTCGTTGCGCAT 170976
      170976
170982 AGTGCCGTTGAATACGGGAGTAGTTGACTGTGATGATGACCTCTTAT
...
171467 TTCTTCATGCTCAAAATAGAGTGTAAACAGATTGAAAAATATATGCTCTTGACGATCTC 171517
      171517
171527 GGTATTCGTGTTTTTAATGCTTCAATGCTTTCAATATGCTGATC
...
171588 ACCAAGACTTACCATTTCTAACCACTCTGCTGTATCTCTATTTCCGTAGCGATTCGC 171638
      171638
171648 TACCACTTATTTAAAGATTAAGCGTCTTTTCAGAGTCAAAATTCACACCACTTATTT 171668
      171668
171708 GATTTATTAACAACCTT
...
172022 AAAGAGCAATCAAGAGCAGTAAGAAAAAGGTGCAAGAGCAGTACGGCAAAAGCAACT 172072
      172072
172082 AGCGGGCTAATCAATGCAATGATGATTTTTTTCTTAAGCTGCTCA
...
172149 CCGCGCGCTTTCTGTTTATGCTCTGATGACAACTTCTGGCGGTGAGCAAC 172199
      172199
172209 AACCTTCTGCTTAATACGCATGTGATTAATTCGCGCCAAAGGTTG
...
172449 GGAATGATGATACAGCGCGCAAAATTCGCTGAAGAAAGCCCAACAGCATGCGATGAG 172499
      172499
172509 GCACGACGACAAACACACGAGGAGTAGACTTTTGTACAGGCCATCACACCATATTT 172527
      172527
172569 TCACATAGTGTGACGCGCTACTGCGCAGACAGAACCGAAGCATGCTTGCCAGACCA 172614
      172614
172629 TCGCTTAGAATGCTCGGCCCATATACGGGTCCATATTTGCTCGCGTCAATCCGCGGAC 172682
      172682
172689 GCTTGAAGATGACCTAAGTTTTCCGCCAACGAAATCAACCGCCAGCGGCG
...
      172689
```

[illegible]

```
...
162998 AGGCATTGATCACCCTGTTCGCCAACCCCTGAAAACCCCTGAGACGACAGAGGCGCTGGCT |-----|
163058 TTTATTCCCGTAAGGCGGTATACCCGCACACCTGAAGTAATCTCATTTTCGGATGTTAAC |-----|
163118 GTAGTGGCGA 163072
...
163189 AAGGACAGTGCCTGAAGGTCCGTGTGTGTGAAGTCACAGAAAGAGCGCTGACCGC |-----|
163249 GGTGAATTGTTGATGTGTACGGCAGGGTGCACGCATTCTATCGTCCGGATAAAGAGAA |-----|
163309 ATTGTCTGTGGCGGATACATCCTGACGGGCGG 163289
...
163369 GAGATCATTCACCATGCCATGAGACATGTGTCTACACCAGCGGAATATCAGACGATTATC |-----|
163429 AAAACCGTGTGAAGAACCCGCGACAGTGATGATACGTCACACCATTCGTAAGACCTGGCGTAA |-----|
163489 GTTGATCTTCTCTATGTGAAGAAATC 163459
...
163604 TGGTCAAAAGGGGTACTGCGTCTGACGGGGTTACTGAAACCGAGGAACTGACGATTATCA |-----|
163664 GACTTGTTCGGAGACCATCCGTACTGTTAGGCCAGCGTGTGCGGAAAGTTACACGCCGC |-----|
163724 GTGAGATGGCGCGGCGCATCGTTTCACTACTCCGTAAGTTGTTAAACGATCCGTTCA |-----|
163784 AAGTCCCGGAAGGTGAGGGGAGCGGTTATCGTGTGATGACCTTGCCAGAAATGATGAATCTC |-----|
163844 TGGCACCACAGATTTA 163805
...
163990 ATCGAGAGACTACCGGAACTGATGCACGATCCGATCCAAATTTACCGCTGACGACAGAA |-----|
164050 AGAAATCGGGTGTGATGTGCTGTGATGCCGTGATAAAAATGCTG 164040
...
164167 TACGTTACAGAAATGAGAAAAAACTGAAGATATGAAAAATGACCGGCTTAACGTTGTAC |-----|
164227 CGGAGAGAAAAATTAAGCCGCGATACCTTCTGCACAGAGGCTCC 164217
...

164429 AGGTGCTGAAGCTGTTTCAGAGAAATATCCGTAAAACTGGCGCAAAAGTGGACGACGTA |-----|
164489 ACAACGCTTATATGCGGAGAACTCTTCCACGGGAGGCGGAAACGACCTGAACGTGA |-----|
164549 TGAAGAGCGCTACGTTTCAGCCACGCTGCTAAATTACTGCGGACTACAAAAATTGGCAGG |-----|
164609 CC 164555
...
164741 TACGTAACAGTGGCAAAACAGGCGACAGTATGACCGCTGTGGCAGGATTTATGACGATATGC |-----|
164801 TGGCCCGTCCGCTGAGAGCTTATCCGTGAGGCGGACCTTGAAGAGAACGCTGTGTGATG |-----|
164861 CCTGGCAGAACGCTTACCGTTACTACGTTCCCTGAAAGGTCAAGATGTGACGGTGTGG |-----|
164921 TGTC 164869
...
164968 AGCAGGACGCGCATGGGCGGTGCATCCCGGGCAGAGTCTCCGTCCACTCAGCGCATACG |-----|
165028 GACCTGAGCGCAATGCTGTATCCGCATTCGCAGAAAGAGTGGGTAAAGCCTTCTGTGAA |-----|
165088 CTGTTGCAGATTAATCCCGACAGAGATTACTGGCAGGATTTACCCGATGACAGACCGGAT |-----|
165148 ACCATGGCGGAGATTGACAGACGCGCAGAGAACCTGTTGAACCATTCGCGAACTT |-----|
165208 GTCGAACGCCCTGTACCGATGCAATGATGACAGACCGGTACTTCACCACCAAAAAGAAC |-----|
165434 TCCGTGACGTGCAGACGCGGCTGATGAACCTGAAAGCGGAGCAGGAAAGCAGCGTA |-----|
165494 AACTGAAAGGCTGATTAATCTTATCCGCCCTGGCTGTGTGTAAGA |-----|
...
165651 AACCGGCTGTTTAACATGGGTGACCTTGAAGGCCAGCAGAGAAAGAAATGATCGCTTGT |-----|
165711 CTCACGTGGCAAGGGGGAGTGAAGGCCAGAGTATCGGTGATGAATTCGTTCCCTTAA |-----|
165771 CCTTGTGAGAGATCCAAACGGTGGGTTGAAGAACGCTGTGCTTTCTGTGCTATTAAGA |-----|
165831 CGCCCGTATGCCCGTTTGTTCACGCCAGACGCGCGCTCTTGTGCCAAAAACATGACGGT |-----|
165864
```

```
160015 GAAGAAACCCGCATGCGTATCTGTAACCCGTGCAGCAGACGACGACGCCCGGCAACCA      |-----|
160075 GAATCTGTTGATCCCGGTGGCGCAGCAGCGGTGAATC      160054
...
160143 GCAGTATCTGATGCGGATATGGCAACAGAGGGTGACGCATCCGCAGCTCATCGCCGGGA      |-----|
160203 AGGTGCCAGCCAGTTGTTGAATGAACCTGACAGACAGACGACGACGCGCTGTATGAA      |-----|
160263 TGAGCTGAAGCGGAAGCCGCTCTGAACCTGCTTGAG      160244
...
160345 GAACAACAGTTTCAAGGCATACGAGAAGTCAATCGCCCAACAAGAGAAAGTACGCGCGGA      |-----|
160405 GCACAGTCACAGCCTGAAATATGCGAGAGATGGTAAAGGAGCATTT      160395
...
160526 AACGCATTGAGAGAGTGGCGCCGACAAACCGTTATTTCAGCGATGAGAAAGCGCCAGG      |-----|
160586 GGAGTGATGTTTCCGTAATGCGCCCGCCACCGCCTGAAACCGTC      160576
...
160872 TTCACAGATGCGCTGAGGTATGCGCGGTCCGACACACAGAGTTAACTGACCTCCACG      |-----|
160932 CCTGCCCGCTCCCGCATATTATCCCGGACAGGATATCCGTTACCGGAGAAAGTGGC      |-----|
160992 GCGTACGCCGGAATGAAATCAAGCCGAGCGTGTGCTGCTTTTACCAACACCGGTGAGGT      |-----|
161052 TAAAGGCCAGAGTTTCCAGAAAGACAAGCTCCCGCACCGGAAACCGCGTGGTCCCA      |-----|
161112 GGGGGAACACTCGAGGGTGACATGTTGTCGTGCTGTGCTGCGCTCACCG      161095 161105
...
161449 AAGGGCGCGGCAAAACACCGTTATGCACAGCTTGCTGAACACTGGGTAACACGCTGAAA      |-----|
161509 GACAGGGATTATGCCACGACGCGCTGAGCAAACTGTTCCCGAAGC      161499
...
161560 TACGCAAAACTGGCGAAGAGGTGCTGATGCTGATACCTTGCGCAATGATAGCGCTGAT      |-----|
161620 CGTAGGATATTTCGCGGGAAGACGAAACACATACGGCAGGCTGGGGGAGAGCATAAA      |-----|
161623

161679 AAAAGTACGACACAGTGTATCCGAAATGCTGAACGGAACGTCAGCGCGCAACGCTCC      |-----|
161739 ACAATGATGAAAGCAGAAATGCCCTCCCGTTACGCGGATACCTGG      161729
...
161823 GGCCTTCTGCTTATCGGGTGTATCGGGTGTGTATCAGCGCGCAGGAGGAAACGCTTACGA      |-----|
161883 TCCGCCACAGAAACTTTATTTCACTCGCGCAATTAAGACAAATTAAGGGG      161873
...
162014 CGAAAGGTATGAAAAAACAGCAGCAGCTACCGGATGACAAAAATCCGCTTTGAGCTTACC      |-----|
162074 GGAATACCCGCGCAGTGTATTTTATCGCTTACGTTAAAAACAATAATGCGGCTGAGAG      |-----|
162134 GTGCTTTAAGTCAGCCAGTGA      162088 162100
...
162191 TTGTTTCGTATGTGAAGAGATGCGGGAGATTTGCGGTGAGAGAGCGGCAACGCGCA      |-----|
162251 ACCGCGACCGTACCGGACCAAGACGCGCGCAAGGGGAATGTTTCAACCGGAGCATTCAGT      |-----|
162311 ATGCGTTTGTTTCCGTTGTGTGAGTTTGTATACGTGGAAGTCCGCTGCTACAGG      |-----|
162371 CTGATTTGAACCGGGCTTATGACTCGCTGCATGACCTTGCGGAAGTACTGAATGTACCA      162362
162431 CAAAAGCGCTTTCCTGGAACGCTGCTTGGCTTGCGCATTTGTGTCCTGCTGTAAGGTA      |-----|
162491 AGCGCGCGCACACTAT      162452
...
162549 ACCGGTGCCTGCGCGCACGAATGTTCCATTTCTGTGATTAATTTTGTGCTGATATGA      |-----|
162609 CGTTTCCAATGACGCGGAAATTTACGTCAGGTGGGACTTTTATGACGGAAGCACAGCGTGT      |-----|
162669 CAGCGCATATTTTAAAGCGCGCAGGTATGTTGATGCTGAATATCCGTACTCAGAGAGT      162640
162729 TTACGACGCTTTTAAAGGTGTGATTCAGGCCATTAAAAACAGTACATGCGCGCTGCTTC      |-----|
162789 AGGCGCTTCTGATAGAGTGGCTTCAAAACCGTACTGCTCAACGATGTTGAATGGCGGC      162733
162849 ACGTGCCTTTGAGCGTTATGTTCAAGATTAAGCGCGGTATGCTGCGCTGAGAAATGATTA      |-----|
162860 TCTGCTC      162860
```

```
...
156442 AACTCTGTGGCAGACAGCAGCTGTGGCAAAAGCTAAATCGATGACAGGGGGCTTTACG |-----|
156492
156502 AAACAGCAGGGGGAGCGCTCTCTGTGTGGCGCTGTGGGGTGGCGCAGGTGCATATGCG |-----|
156529
156562 ATGCATTAAGGCTGGCGGTACCGGAGAGAGTGGCGGTCTAAACACCTTGGCACCAGCGCT |-----|
156616
156622 AATGCCATTACAGCATCAGGCTCAGAACTGGAGACTGTAAATTATGAGATAT
...
156687 CTCGCAAGGGCTGGTTATTCAGAGACAGACCGCCACAGGGGTGACTGGACAGCTCAGCA |-----|
156737 156746
156747 GCGCTGACAGCGCAACAATACGAGAGCTGGCGAACCAGATATTTGCCCGCGTGGGCAAGG |-----|
156751 156768
156807 CTGGCTGACCTTGGGGA
...
156977 TGCCCTGGCAATTGTGGCGCGCAAAAATGSTATCCGTGAAGCCGACAGAGATCGCGAT |-----|
157027
157037 GAATATTTGACGGGGGCTTCTGCACCGGCAAGACAGAACTGACT
...
157142 TGCAATCAGTCCGGCGGTGAGGCATTACAGGGAAATAGTGACGACAGACAGCGTGA |-----|
157192
157202 AGAAATTTAGAGCTGCGAACAACAGATGGCGGCAACAGAAAGCG
...
157493 ATAACGAGCGGATTTTGAGTTTGCGCAGTCTCAGTTTGAACATATATAAAGCTTGATC |-----|
157543
157553 AGCGGAACCTTTGATTACAGAGCCAAAGTTGACGACCCTAATTATTCACCTGAAGGAGAGG |-----|
157582
157613 AGTTTACGGCTAACAGAAATTACCGGATGCGTCACTGGGTATGAGACAGACAGAGACTCC |-----|
157618 157664
157673 AGTTGAGAAAATACAAACAGCAGCGCTTGATATAAGATATGATTTGCCATTAGCCAGC |-----|
157691 157708
157733 CACTAATGGAAGCGCTTTGGAAAAAGCAATTGAGGCTGGCGATCAGAGGCTGCAACGCCGTC |-----|
157743 157768
157793 TGTTCGGGAGAGCTGCCAAGGAGCATCCATT
...
157861 GGTCAAGCGGTGATCAACCTGCAAAAATCTTTGTGTAAAGCCGACATGGCGATGAT |-----|
```

```
157921 TCCTGAATACCCCGGAAAAATCTCGATGTGCTTCCGGCGTTTCCCGCGAACTACAA |-----|
157981 CAGCGTATTTGGCATGCTGATTCAACCGGGAGCAAAACGATAAAAAGGCCAGGATTTGC |-----|
158016
158041 AGTATCGTACCAGCGCAGCAGGAAGGTAAGTACTATTATTTGGCTTGTATCTCATTACAGC |-----|
158098
158101 GATGGCTCCACCCCGCATTAACCTGTAAACAGAAATACGCGCAGTGGCACCCCTGATATCAA |-----|
158103
158161 ACCGTGCTGGCGATACCCGTTGATTAAGGCTATGCTCAGGTACAGGATCCGACGAATTT |-----|
158169 158191
158221 GCAGAGATATCGAAAAAATTATGTTA
...
158685 CCATTGTGCCGGGTTTTTTATGAGAGTCTGTATGCTTATTCAGAGGAACAGCGCTGTA |-----|
158735
158745 GCGCAACTCGGTACACAGATCGTAACAGCTTGAACATTTCAGCAACCCGGCGAACTGA
...
158805 CAGCTATGAAGCATTTTTTCTGTATCCGAATCCCTGGAGAGATTAACAGTACGTGCTCAG |-----|
158834 158856
158865 CTTGGCGATGTATTTGCCAACATGGGTAAAGTTTCCGCCAGTCCGTCCGGGGAGACAGG |-----|
158869
...
159148 GCTGTCAGTATGTGGCCAAAGCCACTGATGATGCTGACCGGTAAGATGTCCGCGGT |-----|
159198
159208 GCAGTTCGTGGCGTGAATACCGCATGACCGAAGGCTATATGATT
...
159535 GTTCTTAAATGCTGATTACTCTGAGCTGGCGCAGTCAACCGAAATTTACAGACAGCGTTTG |-----|
159585
159595 TTCAATTGACCGCCGACCCGACAGCAGCAGCTTACTATGCGCGAAGAAATGATCTTGCA |-----|
159601 159632
159655 AAGAGCGGTGTGCGCATGAGTGAAGTGGCGCGCAG
...
159775 ACAAGCAAGACCGTTAAAGCGGCAATGTTCAGAAATGCCACGGAACAGGGGGCGCATTAAT |-----|
159825
159835 GCGGCGAGGGGCGGCTATTCAACGCTATCAGGAAAAACAGCGCATTTGCTGAGACCGCGGA
...
159895 ATGGGTCTGTACCGTGGAGGGCGTGGCTGACGCAACGATTCGAAGTGCAGCCTTTGCT |-----|
159930
159955 GCTGCATGGGGGCTCCATTCCGTTGCGGTTTGCCGATATGTTGACAGAGCTCAGCCGCA |-----|
160001
```

```
...
152848 TGTGTCAAAAAAGCGGAGTTGATGTATCACTGCGGGGATCGGGATTGTGCGCGAA |-----|
152908 AAGACAGTGAAGCAGCGCTGTAACTCAAGTTTCAATTTCTGCGTC 152898
...
152988 AACACAGCCTATCCGTTTGCGGTATCAGTGGCAAGTACGATCCGAAGCAGATGATT |-----|
153048 TATGACGGGGTATTGAACACTGCTGTCGCGGAAGTTGTGCGCG 153038
...
153254 TGTAGACTGTTGATCCGCAATGATAAATTGAGTTTATGACCAAGATGGACGCTGGC |-----|
153314 TGTTCGCATTAGCAGATTACGCTGATVCAGAGAGTGAATTGGAATTAGCGTTTGCCTTT 153304
153374 TATAACAGAAATTAACGTTAATGTTCAGGGCGCTGCTGACTCCAGTATTTTCTCGACAGA |-----|
153434 TTTACAGC 153386
...
153679 ATAACTCCAGGTTTACCGCCAGTGTCTGATCAACAAGGACACCATTACGGCGACTGAA |-----|
153739 AAGTCTTTTAAACATCAATCCCGTCAGGCAAAAGTTGTTTTT 153729
...
153896 CGACGGTTTANGCTTTCGCAATATGTTACTCCACCTTTCAGATATGTTATTCGCAATGT |-----|
153956 ATAAACCGGACGTGAGATGATTTATCATGGCGAAATGATGCTGCT 153946
...
154578 CAGTGAATATGAGCAAGATAAATAATCCACAGGGCATTAATATCAGGTATTCGCTACGA |-----|
154638 GATAACGGATGATTTTGGCGTTATCACCTCTTTTACATGGAAGCGCTTCTCTCACTAATC |-----|
154698 ACAGACATTTAT 154654
...
154808 CTTTCATGGCGGTTAATCAGTACATGAGTTTATATGCGATGGCAGGATGGCGCTTCTTA |-----|
154868 AAGTTACGCTGACCTGAAAAATTAAGACAATATTAACAGTAGTGGCGATTTTCTGAAA |-----|
154873 154873
154928 G
...

154979 ACCGTAAATGAGACTGTTTACACTGATGTGGCTTACGAAGTTCCGGCTCTGCGACTGCG |-----|
155039 GCACGAGTGGCGTTACTGCTGGCATTTGGCCTGAATTTCTGACCTGTATCCGTAACCTTT |-----|
155099 TACT 155047
...
155182 GCTGGAGGTGAACCTGAAGGCTGATGTGAAAGGCCCGAGTCACTTAACGTTAAACC |-----|
155242 CGAGGCCCTAACACTTTCGTACCTTAAGCAAGTTAGAAAGTTTACGCCCTCTGTAAAGGA |-----|
155302 GTCAAGCGCTATGTCGCAAAAAATCGCTTATCACCGTCACAAATTTGCAAGCGTTATCTT |-----|
155362 CACCATCTGATGTTGCAAGGTTCACTGTGTGAGTTCCGGCTGAATTTGTGGGAGCGGA 155306 155323
155422 GTTTCGGCGCTTCTTACAGTGTAAACAGTAGGAAAAACGGCGACGAGAGTCCCGG |-----|
155482 TC 155428
...
155486 CTGTTGCTGAAGTTACGCCATATGGCACCCGTTTCAAGTGAGAGAAATGAACGATAAAA |-----|
155546 TTCTCTGTATATATGACAGCGTGTGTGAGAAAATTCGCCAACCTGAATTTATGATGAG |-----|
155606 TTAAGAGCGCTGCTGCTTAAAAAGCAGCGCTTTTGTTTGAGGACCTGATGGCTTTTGG |-----|
155666 TGTGCGCTTCTGTGCTCAGTG 155611 155631
...
155690 ATGTTATCCCTTATGTTCTGTGTGCTGGGCTGTGTACGGGAGTAACGAGCTTGAGC |-----|
155750 GTTACCTGGCGGAGTGAAGACATTAAACCATCTTGCGTGGGAGCGCTGGGCTGATTTCC |-----|
155810 ATACGGCAAGCGGGGATTTATCCGCTGGGAAACGACTGGGCTTTGAGCGAATGGCGG |-----|
155870 ATGATGAGATGGCTTCATGCTTTCAGC 155843
...
156209 AACAGTTTGGCGAAGCAGCAGAGAGCTTGGGGTATCTTGTCTCACTCCGGCGTTAGC |-----|
156269 CATCCAGTGTGCTGTTTAAATCGGTATGATGAGTGAACAGAGTGAACAGGTAACCGGGC |-----|
156329 AGATT 156278
```

```
...
149577 ATGCAGAGTCTGTGCGGTTGCGGAAGCATGACTGTGAAAACGGGCACTCCGTTACA |-----|
149637 GCACAACTGCGACTGGACCGCGCAGACATTTCTGGAGTCAGATGAGA 149627
...
150012 AATAATCTGTGATAGTCAGTTGGCCATTTAATGCGCAGATATGATATCAGAGCAGCGCGTTT |-----|
150062
150072 TATTACAGTGAATCTGGAAAAATTGAAAGGGGTACATGATGTCCGCTTGCCAGTGAATG |-----|
150112
150132 GGAATTCACACGAAGACAGTGAATAGCTTTCTGTAT
...
150522 CGCTGTACTGGCGGAGGAGTGCATTTCGGTTTGTTGTTATTAACACCAGACGTCAGAT |-----|
150572
150582 ATCCGGGGTTGAGAAAAACATTACAAATTAAAGCAAAAACAATTTTTCATTACCGTGACGA |-----|
150637
150642 TTTCTGGTTTGCATGCGCGGATGTGTGATGTATCCGACGATCCGATCGCTCAGAGACCC |-----|
150642 150688
150702 CCACGGCGGTATTTACTACACTGACGGCGTTTCTTAAGTGAAGGATGCGACTATTTGC |-----|
150728
150762 CACAAAAGGGGACGGGAATCACCCGACATCATCTGATTCGTGGGGATCCCGCGCGAC |-----|
150797
150822 GACAGCTCTGTGTCTGTACTGTTCAGAGGGGTGATGTTTCTGACGATTAACCCGGAATGA |-----|
150824 150865
150882 TGATGAACCCCGGTTTATACGGAACCTTGTCTCAGATTATGTTGAAGAAGGTCCGCC
150942 AGTCCGGCGTCTGTGAGGTACACATCCGTAATCCGGGAACCTCGGTAACAACGTGAGCT |-----|
150948 150996
151002 GGCCTCCGGTGCATTCGAGAATGCCAGTAAACGTGCGCGGATTTATCGCTCTGCATTC |-----|
151034 151049
151062 AGTTGAGGGGAGGCGGATTTTACTTGTGCGTGAACCTGGAT
...
151111 GTGCTCAGTTACACGGAACAAATACCGCGCAAAACCTTGCGCCTTCCCTGGCGCATGG |-----|
151161
151171 GATTACCTGCGCGCGCAGAGAATATGACAGGCTTTGCCGTGATGG
...
151248 TGAAGTGATGTTTTTGGAGCGTATCTGCCGTATGATGAGCGCGGAAGTGAATCGTCAAC |-----|

151298
151308 GACGGCAGAACATATTTAGTCTATCTGTCCGCTGGGAACGTCACCTGTGTGGCAGCAAA |-----|
151344 151359
151368 GGGGAGCCCTTATTTGTTTCACTGGGGTATCAACCGTCACAAATTTCTG
...
151489 GCAGGAACAATAGCGCTGTGTCTGTGTATGCMAACGGTAAATGTCCGCTGCGCAGGAA |-----|
151539
151549 CAGATTTGTTTACCGGAACAGTGGCAGAGTCAGTTTATCCGGCCT
...
151831 TTTTCATTACCTGAAGAACCCTTTTTCCTGTATCAGGTGGAATATCTCCGCGCTGAG |-----|
151881
151891 CGGCTGGGGATCACCATTTATGCTGATGATGTTCCTGTGATTCATT
...
152105 ATCCACTGTGTGGGAATATGAAATTTGCAACCGGGGACGCTGTGACGACGACACGC |-----|
152155
152165 AATTGTGACCCGTGAGAACTGTTGTCCTTAACTCGCCGCGCTCAGACATCGGGTGG |-----|
152206
152225 TGCTCATATGCACTGAATCCGGATTTGAGATTGACAGTACTTTAATGACTGTTGATTT
152285 TCCCAAAAACCGCTGAATTTTAAAGGACCGGTGATTTGGCTCGGTTCTTCTTGAATG |-----|
152309
152345 GGATATGCTTATTTATCGCG
...
152382 ATCTGGCGGGGTAGGAGATGATGACTTCTGATGCAAGTCTGTGTTCCACGACGCGGGG |-----|
152432
152442 CAGGTTTACGGCGATCCGGTTGACCCGTGGCTGTGCGGATTTTACTGATACGTTTGT |-----|
152451
152502 AAGCC
152559 CAGATCGGCGTGAAGGCCATCATTTGACCAGATCCCGCATGAGGCTGCAAAATTCGCGGTT |-----|
152609
152619 GTGTCCGAGCTGCGTAAAGAAATTAATAAACGCGAGGGGCAAGCTG
...
152667 AAGGATGCTGCATTTAAGACAACCAAGTTGTAGGACTCTCAGGAAAGAACGACAGAGA |-----|
152717 152726
152727 ACGATTGTTGTTAATAAACCAGATTAGCACACTGATTTGTTCAACCACTGAATCGCTT |-----|
152731 152780
152787 AATGAGCTCGACAAGCGATCATTAATAATTGATTAAGAAGAGCGGTGAAG
```



```
146511 |-----|
CGCCAGTCGTTGGTGTGCCGCGGAGGTAAGTTGTTGACGATGACGAGAAAGAAAT
146517
146571 CG
...
146576 TTCCGTATGACAAAGATGATGAGAGTGCCTTATCATCATCCGTAATCTCTGACATTAATC
146626
146636 ATGAGCTTTTGTGCGGTGCGGATGTGATATCACTCCGGACGCTCCCTGCGATTGGCT
146675 |-----|
146696 GTTCTCCGGTGAGGTGAGCGTGTATTATGCCGTGCTCCGAAGCCGACGACCCAGAC
146728 |-----|
146756 TACCGATGAACCTCCTCAATGGCCTGA
...
146850 ACGCGACAGATATTTTCTGTGACGTTTCTGAGGGATCCGTCGGGCATATCGTCATAC
146900 |-----|
146910 ACTGGGACCAAGCCCGTACTCTTTCATACCGCAACCTGTACGACAGAGGTTTTCG
146913 |-----|
146970 ATGACGAGATTACGAATATCATGGAGCTGTGATACACAACGTGTGACCCGATGATG
146972 |-----|
146975
147030 G
...
147091 TGGCAGACCGGATCTGCGCAACCTTGAAACAATAAGTTGTTCTTCCGCGCCGCTC
147141 |-----|
147151 AGTTTTCGCCGATGTGTAATACAATTCTTGACGTGATATGCT
...
147247 AGTATCTGAGTGGCCACAATGANGGTAATCTGAAATGTTTATCAGCAAGCAGCTGT
147297 |-----|
147307 CCCCAGGCTAATTCGCTGTTTCTGCTCTGACAAAGAGATAAG
...
147355 TTGATGCACTGTAAAGCCGATACCGAGGAGTGTATGTTCTGACGACGAGCATGATA
147405 |-----|
147415 CGCCAGTTCACATGAGAGAGCTTATGTTAAACCCACTGATGAGTG
...
147587 CCGTAAAAAAGTGTAAACGAGGTGAGTGTGATGTTGTTTCGAGGAGCGCTGAAA
147637 |-----|
147647 TCTCCTGATGGTGAAGCATATACGAGACAATAATTACCCGTGACGCGCTGACAGTTCA
|-----|
```

```
147707 147661 147691
CGGATGCGCTCAGCGGCACCAGATGCTGCGCAGTGACCCGTGAAGTGATATTACGGA
147767 |-----|
ATGAGATGCAATCCGGGGAGTATGCGGTTTCGGTGTGACGGTGAAGGAGACCTGCTGC
147769
147827 TACGACGTGTGCTATTGAGGGGACCGAAAGTACGTTGACGCTCAATATGCTGTTACGC
147865 |-----|
147887 CGCAGTCTTGTGAGTTAGCATTACCCGAGAAC
...
147928 AGATTTCCGGCAGATACAGAATATATGTGCTGATGACCTTGGCACTATTGCGGCTGAA
147978 |-----|
147988 TGAAGACACGCGGACAAAAAACACTCAGGCCACACAGTCAAAAAAGATGACAGCCAG
147997 |-----|
148048 TCGCA
...
148061 TCTGACAGTGCAGAACAGGCAACACAGAGGGGCTGAAACCCGACAAAAAGCCTGAT
148111 |-----|
148121 GCCACTGAGGCTGCGACCCGTGCTAGTCAACAGCAGCGGGGAATGACAG
...
148215 ATGCAAAAAAGCCAGACAGCATGCTGAAAGCCAGGCTGCGCAGAGACCGCGAG
148265 |-----|
148275 AGATCCTTAAGCGGGCAGAGGCTGCCACTGTCTGATGCTGAAGAGGC
...
148407 CAGTTCAGTGGGCCCTCAAGGGCCCGCAGGCGCAAAAGTGAAGGCTGTCGTTGAG
148457 |-----|
148467 CTCAGGGGCTGTAGGGCTGCTGCTGCTGCGCGTGTGAGAGAGGGCGA
...
148761 GAGGAGACCGGACCTCAGGGACCTCGTGAAGACCAAGTCCGACAGCAGCCTGCAA
148811 |-----|
148821 ATGTGGCTGATGCAACGAGGCGCAGAAAGGAATTGTGCATTTAAGCAGCCGACAGCA
148834 |-----|
148881 GTGATGATG
...
149231 GCAGGCCACAGGAGCCGAAGAGAGAAACAGGTGCGGAGGCCGCGTGGGCGACCGGA
149281 |-----|
149291 CCTCAGGGGCGGAAGGCGACCCGGGGGAGACGCAAAATACGGTTCCGTGCGGCCGGGA
149306 |-----|
149351 AACATTATTGA
```

```
143099 GGGGTTTTTGTTCGAGAGTAAAGCGTTCCGGAGCCGGCTTATTCGGGGGCGAG 143157
143122
143159 CGATTAAGCCTTATTACTCAACCAATTCGATCTGTCCGATMAACAGACCATGCG
143162
...
143265 AAAAAATGCTGGAAGACTTGGGAGTGGGATATTGACATGACGACGACAGACGATCTGTA
143315
143325 CGGAAGATACCAAGGAAAAACATGGGAGTGAAGAACACAGACTAGACGGGCGGATACA
143375
143385 ATGTGGACACCGAGCGCGGATGCCAGTGTGTGAGCAGACGAGGACGTGAAGGAGCCGGAG
143395
143445 CGAAGG
...
143455 GGTCAACCCGCGAGCGTAAACACGTCATTCCCTATGAAGTCTTGAAGCTGAACGTTCCGG
143505
143515 TAAGCAAGGGCCGAACAAGAGCCGCACTTCTGTGGGAGATAGCTGAAGAAAAACG
143575 CAGGCTGAGACTGTGAGCTCTCAGATCCACGAGCCGCTATGAAGCCACACCGTTACC
143590
143635 GGAACACGAAA
...
143809 TCGATGAATGCCGTGCCGTGCTGAAAAAGTGGCAGAGATCCAGATCCGCTTC
143859
143869 GGTTCCTATTCATCGAAGGAGACTCCAGATGACCCCGATGAAAGACAAACGCT
143923
143929 CACTGAAGCTTTCGCTGAAGTGGCCGTGCTACGAGGTTGCTTTCGGTGAAGTCAGTGA
143934
143989 GTCCGTCTCTGACAAACAAGGACAGCAAAACGATATCCGAAAAACGGCGAGAGA
143989
...
144096 TAGGCACACGCGCTCCGTCCGAAACCGGTGATTAATTTGAACGGTTACTTGGCGCTTC
144146
144156 ATTCAAGAGCAGAGCGGATTAATGCGCGGTATGAGCATGCTGAATTAAGCGCGCTTCTGG
144169
144216 AGAAGCTGGGTAACTTAAGAGGAGAACTGAAGTAATGAGGAGCTGAACATCAGCCCA
144255
144276 GGCGAATAAGCTGTATCAGTGGCGCTTTTACCGCTGCCAACCGCAACCGCTCGATGAT
144297
144336 AATATCTCTACTGAACAGACAGAGAGCGCCAAAGCGTTTCCGGACAAAGAAAAGCAC
144360
144396 GAAGCAGACACAGCGCGGTG
144297
...
144444 CAACAGCGCGGTGATGAAGTACCTTCAGCATCATGACAAACTCTCAAAACGTCGAC
144494 144501
144504 GATGGAGATGAGCGCTGTTGAAGTCTGTGTGAGGATCTCAGCCATGCTGACTTCTCCT
144506
144564 GAAAAATCAATCAGGAGCGTCACCTGCTGATGACAGCGGACGTATGATCAGCAGCGCAC
144578
144624 GAAGTTAACTGAGCATCCTCAGCCAGAACGCTTCTGGGGAGCTTAATGACCTGCA
144651
144684 GGACCACTGTGCGATAGTCATCTTGTGTGAGCTGTGTGATTTTGTGCTGAGCAGCAC
144736
144744 TATCTGCCGACAGCGGAGACACCTGAAATTCAAAAAAATCAATGATCAGAGATCTGCGC
144790
144804 TCCGACATGACCGTCACTTTTGTGGCGGTGATCCGACAAAG
...
144979 TATTACGTCTGTACGTCAGCGCGCGCTCAGTGAATGACTGTAACCTCGAGCGTCCGCT
145029
145039 AAGCACTGGAACCAAGATGATGTTGTCGCCGTGAACCGTGCAAAAG
...
145178 GGTCAAGGTTCTGTATCAGAGAAATTAACCTGACGCAACCAAGAAAGAGGTGCGCTG
145228
145238 CAACCAATATTGACCGCGCCCATGTTACTGGGGGCTCAGCGCTGCAAAATGCTTACGGTC
145275
145298 AGAAGCGCGGCGTCACTTCAACATGTTGAGA
145275
...
146238 TCTGTTAAGGGCAGAGAGTGAATTTAAGTAAGTCACTCAGCAGCGCTGCGAGCGCTG
146288 146291
146298 TGTGAGCAGAAAGACTGATATTTCACAAAGACCCGAAAAAGAGCGCTG
...
146391 GCGGAGACTGAACAGATGGCGGAAGTGAATTTTTTACCCTATGCTCGCTCATAT
146441
146451 CAGCGGTCACTGAACATTATGATGACGATGCTCTGTCAATGGCTGCCGTGCAATTCAG
```

```
140429 GCGAGCAGGTGGCTCACTGTTGCGGC
...
140577 TCTATCCGACAGTTATATCTACAAAGAACAGCATCTTTGACCAAGCATATGACGACGATTA
140627
140630
140637 GCGCCCGCCTTGCTGGCTGCTGACACCCGTCAGAGCAAACTGTCTGTGACCGAAGAAATGA
140697
140698 AAACGCTTCTGAATAATGGAATATCAGGATCCGCTGGTCAGGACATTTATCGAAA
...
140893 CAGAACCAACACACTGGATGGCTCACTGATGAAAAATGAACCTAACACCATGGCGCAGAAA
140943
140953
140953
140953
140984
140997
141013 ATTGACACCAACCTAAATGGCGTGATGCCGCAACAAAGGCCTGTGCGTATTACGATGGC
141070
141073
141075
141133
141160
141193
141314 ATGGGAGAACAAATCAAGCGGGGCTCACTTGGGTGGAGGTCAAGACGAACAGCCATCCGT
141364
141374 TCGGGCCTGAATTTAAGGTGTCTACTGTACGCCGGAATGAGTTTTT
...
141423 GGACTGGCTGAGCCGGGAGGCTGATTTAAGTACTGCCGATGGCTGATGCTGCGCCGCTG
141473
141483 GATGGATACCGATGAGGCAAAAGCTACATCCCGGGAATGGCTCAG
...
141615 GGAAGAGTATCATGTCATGGATGATCGACAGACAGAACGAATGGCTTCAGCGTGAAACCCGCTG
141665
141675 TGTGCTGCTTTCAGGTGTTTATTTACCGTACATTCGAGCGCTTCCGGTGAATGAATCAG
141710
141735 TAATGACGGGTGGTGGCCTTTGATAAAAAT
...
141822 GGTAAGCCGATATTCGTGAAGCCTGTTTGTGCGGCCACACATTTATTTGTGATGCCCCCTG
141872
141882 TAGTCTCCGACAGGGGATGTTTCCGCTGGTTCCTTCTTGCGGATAC
...
142025 TTTCAGGCCAAACCGCTGATTATGACAGAGATGCCACCGAGTTGTACAGACAGACCTGA
142075
142085 TGGAGCAGATGCAAGCTCCGGATGGCATTTATTAACCTGAATCCGGTCCGAAAAAATCACA
142097
142145 AAAGTGTG
...
142234 CAGGATACCAGTGGAGTGTATTCCGCTATTTCGCGGACAGATTTCAGGTGCCAGCTCAGGC
142284
142294 GTGGCTATCAGTAACCTGTTGGAGCAGGGGGCCACACCCCTTGCGGAATCAGCATTAAC
142345
142354 TACCACTTTCCTGCCAGCAGGTGGAGAACTGTGCTGCTTATCTTCTCGATGACCTG
142414
142418
142454
142460
142474 ATTGTCTCAATGCTGAAGGTGATATAGTGTGAACCTGACCAATGATTTCAAGTTAAAT
142534
142548
142569
142594 AGAATGTCAGAGTATTCAGGGGCTGCCGC
...
142669 GATGTGCCGAGAAACAGAGATTGTTGTGAGCGTATTCGTGCTGGCTGGGAGCGCAGAAA
142719
142729 TCACCGGATGAATGACGCCGGAAGAACAGAGAAAGTAGCGGCACAAACACAGCACTTCAG
142743
142789 CAACAACAGG
142799 CAGAACTCCAGATGCGCGAGATGGCTGGCAGATGGCAAAACTGGAACTGACGCCGCCA
142849
142859 GGGCAGATGCAAGCTGCACACGGGATATATGCCAGTGCACACGGGAATGCCCTGCACAC
142919
142925
142979 TCACGAATATGGAACAGAGAGCAGGAGCTTCTTCAGCAACAGATGCTGTATACGTTACACAC
143002
143039 AGCGGATGAATGAATAATGCTGCTCTGAATAACTCTGCTTCAACTGACACCCGTCATGCTAC
143055
```

137177 TCACCCGCTGGCAGTCTCTGACGTTATCGGAGACTGTGGGTGAAGAAGATGACCG
137197
137237 TGTCTGTGTGTGTG
...
137450 TGCAGAGCTGAAAATGAAGCTCTGCTGCTGATGTGTCCGCTGGTCGTAAGCGCTGCG
137510 GATCAAGCCGACCTGCTCCGGTACCCTGCTGAAGCCACCGGACCTCCGGCGTGGAATA
137514
137904 GGGTGAATGGTTCTTTGGTCTACTGCTGTCTGTTGTTTGAACAGTAAACGCTGT
137954
137964 GAGCATCTCTGTATAAGCAGGGCCAGCGAGTAGAGTAGCAT
...
138092 GTTGAATTTTCATATTTGTTAATTTATTAAATGATGTCAGGTGCGGATGAATCGTCATTG
138152 TATTCCCGGATTAACTATGTCACAGCCCTGACGGGGAACCTTCTCGGGAGTGTGCGG
138212 GGATTAATCAAAACGATGCACACCGGGTTTCTCATTTTTCACGAGATGGGCGCATTTTC
138223
138272 CCGGAGAGCCGCTGTCCGGTGGGTGGAGAAACCGGATAAAA
...
138361 GTGAATCTGAAAAATCAGACGGCTGATTTATGATCACTTTTAAACAGACGACGACAGAA
138421 CGTGACATGCAAGCTGACTGAAAAAAGCTGGAGCAGCATTC
...
138599 AAATCGATCATGACATCTCAGATGAAAGCCCTGCATGACGCGGTTCCGCGCATGAAA
138659 AATGTGCAAAAACCTCTGAAAAAACTGCGAAATTCGCGCAAAACGAAACGATTCGTG
138719 GTTCCCGACTTTTACACCTTCAAAACGCTTTTCTCAGCGAAACACCCACCGCGTAAGAC
138779 ACCGTGATATCGGAAGTA
...
138969 CAAGCTGAACAGGCCCTTGAACCGAATATTGCCGTATGAGTCAATTGAACGCTCATTT
139029 GCTACGCTGGACGCTCTGGCTGAGACAGACCAAAACTGTGCTGACCGGAAAGAAAT
139032 139039
139089 CAAGCGCCGCCAGAGTAAACTCAGAGCTGAACCGATATTCTGACCAACACGCGTGGGG
139090 139139 139147

139149 ---| |----|
CGTTGTTACGCCCTGTCACTGACATCGTGTCTCATTCGCTGCATGAAATGACTAATTCGGGAG
139152 139179
139209 ACTGATGACATTCGCGAAGAAATGAAACCGCATGTGATGAGCCGTCAGAAATGACCGAGG
139269 |----|
CTGAACAGCTCTGTTTCATCATGACTAACTGAGCAATCCCTGCTGGCGGCTCAATCATC
139275
139329 TC
...
139336 AATACAGAAAGAAAAGTGAACCTGTACCTTCAGAAATGCGACCGCGGACGCAAT
139396 GTTCCGAGATGCACAATAAAAATATTATCTGAAAGCGCGCCAG
...
139464 ATATTATCTTCTGACAGCAGCATTTATTCATTCGCAATCTCAAAATGCGGATCGTCGCTC
139524 AGGATAACAGCGCTGCCAGTGAATTTTCCGACAAAATTTGCTGTACCGTTTGTATCATC
139584 TCCCTGACTGGCTGAGAGCTCATTTACCAATCGTTGAACGTCGTAGCGTCGACAGCGGTG
139644 GCTATATCTCTGTTGGTTCACGGCTCGAGTATTCAGGTGCAACCTCATTTCCGTTTACGTA
139704 |----|
CGGTGACAGCGCTGCATATCTCAGAGCAGGCAAAATTTGGCGAAATATCCGGCTAAGG
139764 CG 139710
...
139911 CACAGATTATTAATTCATTTTACGCTGTGTGAGAGATCTTAATTAACAGCGCCAGAG
139971 TGCCGAAAGCGGCTGAAAGCTGTCACGGAATAATGACGTATTT
...
140075 ATCAATAAGAACTGAACAGCGTGAGGAAATGAAGCAGAGTTTCCCTCAACGCGACAG
140135 |----| |----| |----|
GAGCGCTTCTGACGTCCGAGCAGAGGTGTGTTCACTGAGCCGAAGTACGTTGACAGCAGAA
140137 140146 140154
140195 TCATTCTGTTGCCA
...
140249 AAGACTAAAGCTCAGTCTCTGCTGAAGAAATTAATAAGAGTTGCAAGCGGACGCTGATG
140309 AATTATCTGCTGTATGGAACTGCCGATCCGATGAAAGATATGTTTGTGGGCGAGAT
140369 ACTGCCGAAGGGCTGAGACACGAGACCGGCTCATCGCTGATGTTGTCAAAACGAGTAAT
140401

```
133859 GAGACGTGCCACTGATACATCTGGCCACCTGATTCCTCCGTAACACTGACGCCAGAC
133919 TTTACGACACCGGTACACCTGATGATTTTCATGATACGCCCTGATCTCTCTCTAG
133979 CCAGTCGTGTCGTCCGCGACGGGACACTGCGTTTATCCGGATGATGTCGCTGTGTCGACA
134039 ATGTAATATCGTTGACGGGGCAATATGCAAGTTCCGTCAT
...
134138 AGCGGTGCACTCCGCTTCGCAAAATAAGCGAAGCCTGGCGAAGATATCGTTACTG
134198 CGGCGCAGTTCACGATTTTTCAGGTTCCAGCTCTTTACAGCGCTGACGTTTCAGCGCTGGTG
134258 AGCCCACTATCACCGCCCCGGTATCCGCTATGCTGCGAACC
134355 TTGCGCGCCACTGTGAGTCATTCATTCCTGACTTTCAGAACCATACGAATCCGCCGCT
134415 GACGACTTCGGGGGAAAAACGAGTATTTTATGTCATCCTGTTAC
134461 CTCTTTCTCAGGGAGTTTAGTCTCCAGGATTTCCGGGGGCTTCAAAAGTGGCAGCAAAC
134521 AACCTGAACCTGGGAACCTTTGGCGCATTTGATTAACGAATGCAATAGCGTGGCTTCCGC
134581 TTTGCCGGGAATACAGCCTGACAGTGAACGCCGCTTATTTGATGTCAGATGGACACCG
134641 TTACGCTGACGACGATCACCGG
...
134742 TGCGGTGAAGTGAACAGACAGACACTGCGGATGATTAATAAAACCGCCGACGACAGGA
134802 ATGACCGCTGGCGGTGGTGAATACCTATGAGAAAAAATAAGGACGATCTTCTGTAAT
134862 CTGGTTTAAATGAACAGTTCTTATTTGTCACAAATAA
...
134995 TGCGTACTTCTGGCGAGACAGATGACAAAAGGCCGCTGAGAAAATAGCGCCAGTC
135055 AGTCAGCCGGAAGCAACCCGAGAGCGCTCATTTAGCGGCAACT
...
135125 AAAATCCATGTGAGAAATTAATACCTCGTGTCACTGCAGAGGAGTGAAGAGGTCTCTGC
135185 GCTCTGAAGAAGTCCGAGCGCACTGAACAGAAACTTGGCCATTA
...
135308 AAGCGGAGATACCAACCGAGAGAGCGGTGAAGTTCAAGCTGATACACCGGTGCGCATG
135368 CGACTGAACCTCAACCCGATGTCATGATGCTGTAAACGGGAGT
...
135693 AGTAAGCTGATGACGAGAACTCCGCTATGCTTTATCGGCTGCCGTTCTGGCGCTGAT
135753 GCCGAGGTGCTTCTGCGCTGAAATCCCTCGACAGTTTCTGATG
...
135883 AGCTGTATTCTTGCCATGAAGCTGTGCAAGAAAAATGCGACCGGCTTAACGCAATG
135943 AGCTGATTAAGCGCGTGGCATGGGTGAGAAAAACATCAAAAGTCCGCTGACCGAACC
136003 AGAAG
...
136014 ATCGCTCATCTGTCGCTACACATTTGTCGCCGTAAGTGTTCCTCGACGTTTAC
136074 AGACGAATTAATGACAGTGTGCAAAAGTGCTGCGAAGCTATT
...
136540 CTCTTATATGATTAACCATATGCGCAACCATAGTTCCTCATGAACCTGTGCTGG
136600 GGATGGGCTGGCTTGCAATCAGAGTTGGCTAAGTTAAGACAGCGTTTGGCTCAACTA
136660 TAAGGAATATGCTATGCTGCTGAGATGGGAACAACCAATATGCTGCTATGCCAC
136720 TTGC
...
136762 AAGTAAACCCGAATCCGATAGGTCAATCCGATACAGGAAGAGTGGCAGCATGTC
136822 TTTACGAGTACTGGACGAAGGTTTGTCTTAAATCCCGCTAAATAGCGTATGGAAG
136882 AACTCAACAGGCTTGCTGA
...
137057 AATTAACAAGCGGATTTGATGGGTGCTAATGATATCAGATTAACCTCATTAACGCTGTG
137117 AAGAGCTCTGTACTTGTGCAATTTCAATTTTATGTCATGCTGGTGCATTTTCTTT
```

```
130742 TATACACGGGTAAATAAGAGTT
...
130869 TGAAGATGACACATTTCAGTGAAGGTTGACGGGAAAGAAATCTGACCACTCCCTGGAA
130919
130929 TCTGCACACCGTTACTGCAAAAGTGTCTCAGTTGACAGAAATGACTGTC
...
131120 CTGAGCATTTGTTAAAAAAATTTGCGATGGTAATCCCCCTGTGTGGAGGGGCGACTGGT
131170
131180 GAAAAATCCTTGCTTGATTCATTATCGACACGGGTTCCGTGGTA
...
131509 TCCCGCGCGGAGGATTGATGCAATTAAACACTATGATGTTGTCAGGGCGCGTCCCG
131559
131562
131569 TCAGACCTTGGGAACGATAACTCAAAAACGTGAAGAGAGGCTGGCAGC
...
131674 GTTGTGAGTGAAGGGAGCATCTCATCACAGAGGAGCCAGCAACCGTCCGACGACCTCA
131724 131733
131734 GCGCTTCTGTGTCGCGCAGAAATATTACTATGTGATCGCGCTTGCTGCTCAGTCCA
131738
...
131858 AGTTAGGCGCTCGCAGTACGGGTGACACCGGGCGGTGCAGTATGCAAAATATTAACGACATCA
131908
131918 TTCCGGGGGACCAATTGTCTGATGATGTGACAGACATGACCCGTCT
...
131975 AAGCGACCTGTCTAAAGGGGAGTACGGAACCGTGGGGCAGGGCTGCATATATGCGCAAAA
132025
132035 AACTGTCGCCGTTTATACCGCGGAATGCGGGCATTTGCTGTGCTCCGTGTCGTGGTG
132095
132099 GTTCAGCGTTCCACCAACCGGAGCGGATGGCAATATACAGTGAACGGAGTGTGCTCGSAGA
132099
...
132571 CCGGGTATCGGATATTACGGTTGGAATGGCGTGACAGCTCAGCCACTGAGCTCACAG
132621
132631 GACAGGGCGAGCAATTTCAAGTTCATGGGCTCGCCGCGGATTAATTTCCGACCGCTGGCA
132691
132693 ACGGCAATTTTGGCCATGCGGGAAGAGTGGCGCTTAACCGCGGCGATCATGACACGTA
132720
132751 TCAGAGGTGGCGCGTATGCGCTTCGCGTGCAGAAAGCCACAGGCGTCAACAACCTGCTC
132768
132793
132811 TCTTACCTTGGCAGCAGTCAAGAGGAAGCCTGAAGTACAGGATGTTGACGCCAGTGGC
132871
132889 GGCAGGGCAGAAAGTGTCTACGATGCCGAGGGAACCGAGGTAAAGCATGCACTGACCC
132931
132931 AAGGAACCGCGTAA
...
132985 GCTCTGTTACAGAAAGGGGGCAGATTCGTCGCCGTTTAAGTTTCGGAGCCCTGGCT
133035
133045 GCGAACCAATATGTTATAGCGCTTTACTGGCCGCTATCTTCACTGCCACAGGGCTTGGC
133062
133105 CTGACCGGAGACGGGGGAATAACCTGCTG3CAGCGTTCTATCATCCAGACATGCAAAA
133137
133165 GACCTGAATGTGATGATACCAAAATGCGA
...
133203 CCCCAGTTTCTGAGAGTGTGTTTATCTGTAACCTCAGGCTGCCAGATCACTGCTTCCG
133253
133263 ATGGAAGCATAATTAAGCTTTTCTGCTTGTGCGCGGAGAGATAGCCAGCTTCCAGC
133323
133324 AATGCTGANTGTTATACCAAGTCCACCCAGCTTAAGTGTGGCCAGTTCCACTTCTGCA
...
133410 TCCGCTTTGTAAGACCATTTGATGCTCTCAAGCCATCGCTTGTCATCAAGATCGCTGTA
133460
133470 CTCCTGTGTATGCCAGTAATCCGGCTCTTTTATGTCGCTCCGTAAGGCCAGTGACACA
133504
133530 TACTGAGAGCCTTTATGCGTGTGATGATGATGGTGCCAGACGAGCGGCCCAACAGCC
133544
133590 TGCTCCAGCGCATCCACAGCAATGTGTTTCCATTAAGCATGAGACCCGCCACCCGACG
133614
133650 ATGTATCCGGCAACACATCATATGATTAAGCCCATAGACGAAGCCCTGCCATGTGCTG
133677
133710 ACGTAAGTAAATCAGCCACCACAGCTGTGTAAGGTCTGTGCAACAGAACTGACGGTTT
133744
133770 ACGCGTCCGCTGCGCAACGCTTTCGCGTATGATGTCGTACGGAACCTTTTACCCGG
133775
133830 A
...
133830 A
```

```
127312 AGTTGCGCCGGAATTACACCGCTCGCCTTGTTCAGCCGTGGGAGAACTCGAAGGGGCAA
127316
...
127647 TATTTACAGTGAAGAAACACGGGTCAATACAAATCAGGAATACAGATATCGTTACAC
127697
127707 AAAAAAATCACC GGCGGTGTCAACAGTGGCTGACCAGAAAGCTGC
...
127852 CTCAATCAGTGGTGGTGTCTGCCCGGAAACAGCAAGCACTCAATTAGCACTGCAAGCAAG
127902
127912 TAAAGAACCGGAAAAAGCGGAAAAAGCAGCAGAAAGAAAGCAAGCAGAGAGCAGAA
127950
127953
127972 ACAGAAAGATAACTGAAGATTGAAAACTCGCTTAAAGCCCGCAGTTACTGATTA
128032 ACAAGCCCAACAAAGCCGTAAGCCCTTCATCAGAGAAAGAGACCGGACTTACCATGAT
128051
128092 CTCTGCGGAACGCTACAGCTGTCTCAGTGGGATGCCGACATTAACCGGACAACCTGTC
128101
128152 GGCAC
...
128257 CCGCATCGGCGCAGGAAGCAGTAGACAAATCGAATCGAACCATTAACCGCATCGCTGAGC
128307
128317 TGTGGAAGATGCAAGGCGATCAAGGCGAAGTATCAGCAGAAACTTTAAAGACTCGGAA
128332
128377 CAGCAGAAGTAGGCCGATGACGTTCAAGTAAAAACCATTCAGACATGCTCGTTGAG
128397
128437 GCATATGAAAAATCAGA
...
128456 AGTAGCCAGATACTGAACTGTAGTCGACACAGGTCAGAAAAATACACTGGCGATTAAG
128506
128516 AAGGAAAAAGACACGCTATGCTCAACGGGTGTTCTTATGGTTACCCGCGATGGGGTAAAG
128533
128576 ATACTGATGCGTGAATATCCGGCAGGTTCTTAAAGGCTGGGGGCATGGCGGCAATATAC
128607
128636 TATGAGATGTTACATGTCGCGCATTCGTCGCGATTAAAGGAGACTGATCCCGGAAAA
128653
128696 GTAAATACGCTCCACAGTGTCTGTACGATGATGATATGCGGGTGATAGCC
128720

128756 CCCTTTACCGGAAACAATCGCATTCGATGACTGCTGTGATTATTACGTTGGGG
128816
128818 GAGACGTTTCATGCGCGCTTGGCAGGAAACATGGGTGCTTGACACCTGTATAGGTAAAGC
128827
128876
128876 CTTCACAAAGCGAGGGGATTTGTTAAGGCATGCTGATGATGCTGGAGCTG
128876
...
129254 GCTTACGAGTTTGCACAGCTCCCGCAGTGGCTGCTTTTATATGTCGCTGCAAG
129304
129314 CAGCAATGGCGCTAGGGCGCTCGTGCATTTGGCGTTGAGCTGGAGACGGCGCTTTGAGC
129321 129329 129343 129362
129374 AGACGGTCAGGAGGTTCAGAAATGTATGTCAGTCAGAACGAGATGA
...
129434 ACGTACCGTTATATCTTGCGCCCGCCCTTATAGCTCAGTGGTGAGAGCGGACGCTCAT
129484
129494
129495 129503
129554 TCATCGGGACAGAGCGCCAGCCTTCGAGAGCTGGCTGGCGGGGTTGAGTCTCGATGGC
129566
129614 GSTCCATT
...
130178 ACGTTTCATGACACAGGACGACAGTATATACCACTGCAAGCTGTGGCAGCGCTGAA
130228 130236
130238
130241 130257 130285
130298 TTCACTGGTAAATACATGACCAGAGATGATCCAGAGCAGTTTC
...
130388 CAGAGAGAAATTTGTCAGAGCACTGTCTGAAACTGCTCTGTGTATACGATAGCCCGGA
130438
130448 GACGTGACCTCACTGTGAACTGGGGGCGAAATCAGCAATGTCTTC
...
130502 CGGGAGAGAGATGGTGTCTCAGAGTGGGAGAAATATCTTTAATAATATATCAGCATACG
130552
130562 GGGACTGTGGCGTTATACTGAATTCGCAATCAGAGGGGGCGTTCTGTTCGGGCGGTG
130622
130653
130682 AATACATTATGGGAAGTAATACAGCTGACGCGTTTCTGAACAGAAAGTCAAGTTTTA
130711
```

```
124433 TGAGCTTGTACGGGCTAATTTTCGTGAAGAGTGTGGCTTGTGTATCCGTATCCATTTCATC 124483 124490
--| |-----|
124493 GTCCAGTTTGGCCGCTTGATGAGTTGTATGACATTACCTGCCCGCAGCCGCTATCCCGC 124540
124553 TGACCGTGTCTCTCGACGGTTGTTCAAACTCAACATGAGTTT
...
124598 GCGCAACAGAAATTATCCGAGCTGTGATTTAAGTGAAGAGATGAGCATGAGCGACCTAT 124648
124658 CATTACCCAGCCAAAGCTAAAGATGTCCGTTTGGGGCGGTTA
...
124830 AGCAGAGATGAGCAAAACCTTCTGCTTGCATCCAAAGCAATCAGAAATPACGGCATAG 124880 124889
-----| |-----|
124890 ACGCCATTCTCTCACTACCCATGCCAAGTCACCCACAGAAATCCACGTTAAAGAAC 124913
124950 CCAAGCCGACGAAGCGAGATGACCGTATGTGGCCGATGCTGACAGCATGTTTCGCGTC 124994
125010 AGTGTATATGSCATGTCTACACGGCTGGGCGCCGAAGACTGGAAAGACTGTTTCACTGCC 125035
125070 TGTGGCTTAAGACCAAAAAAACTGGAGCAACGAGTGTGCTGTATCGACGCTGGCGCTG 125123
125130 TCATGCTTGGCGTGCCTACCAAGCAAAATGCGGAAGGCCAGCATGACTGA
...
125278 TGCATGACTATCAATCAATCAATATACGCCAGACACAGACAAGAGATGCTGTGGCAACGCCGCTT 125328
125338 TGGCTTTTGTATGACTGTGATATTGAGTTTGGATTCTGGCTGATTCGGCAGCGAGCGAC 125392
125398 AAAAAATGCTCTGTGTGCGCTCACTGTGCTAAGAGGCCAGACGCGGCTCAATTCTGAGTGG 125434
125458 GTAAGCCACGGTGCATCTGTGAATACCCAC
...
125493 CAGCAATATACAGCCGTGGGTGGAAAAAGCCGTGAGCAGTGCATACAAACGCGACAGAC 125543
125553 GGTAGTTATGCTTGTGCCAGAGATATGTCACTGCGGATGTTTCAGC
...
125667 GACAGGGCTGAGAGAAGAGAAAGCAAGGCTCCATGCTGCTGATTGGCGACCGCTT 125717
125727 CATCAGTCTCGACGAATGTTTACTACTGTATCCAAAGCGCATTTGATGGCGATCGGGCA
|-----|
125787 GGGCGTACAGAGGGCGGCATAGAGCGACAGCAGAGAGATATACCGACATCATCTGGCA 125775
125788 125809 125815
125847 AACTGCAATATACCTTCCACAGCAAGCGTCCAGAAATTAACCGCAAGCAATCCCAAAAGAA 125869
125907 TCTGACGTAAAAACCTTC
...
125989 TGACGCGATGATTTATTCGGGGCTATATTCCTCACGGCGCAGCAAAATCTGCGTGGGA 126039
126049 TTGGCGTCTGTGATAGAGACCGCGCAGATACAGCGCGCGAGCGTTTTATTGTTC 126051
...
126158 CCGCGCGCGTGGGTTTCCCGGTTACGCCAACCTGTGCACAGTTTCACACACACAGCATTTGC 126208 126215
|-----| |-----|
126218 GTCGCGGTGGTGTGATTATTAAGAAACCAACAGAGAGGCGTCAATTATGACAACTCAAAATTT 126220 126253
126218
126278 CTGTTGAATCTCTCTCCCGCATCAACCATAA
...
126827 TATGCCACGACTACGGCTATGAGTTCTGTATATCATCAGGAGGGGAGGCGCTTTTAA 126877
126887 TTAATTCACACCGGGATATAGATATTTATGAGCCTGACGGGCGGAC
...
126936 CCTTCTGCAATGGGAAAGGCTTAAGAACAAAGAGTTGCCGCTTCACGTCAATCCCTACTG 126986
126996 ACAATTGACAACTTAACAACCCAGCTTCGGCTGGGTTTTTTATTG
...
127069 CAATGATGAGCTGATAAATAGCAATGCGATCAATAATGACAAGCATTTGAATTCGCTGAGT 127119
127129 TGTGTGGAAGCGTCATGACAAAGGTGAACAATCCATTGAACGACTGGCGGTTCAGAGGTG 127168
127189 TGATCCGAATCCCCCAATGTTGTTTCGAAAAA
...
127252 TAGAGGCTTACGTTTTTGAAGGCGAACAAGGTAAAGCGGACAGCATTTATTGCTGTGCC 127302
|-----|
```



```
122083 GAGAAATTATTGAGCTAGCTAGAGCGAAATTTGCCGATGATGTTGGAAGAGCGGCCAA
122133
122143 CCATCAGCGCGGACCCGCTCTG6CACAAAAAGGATTTGATTTCTGCTTAACGATTAATAC
122192
122203 CTACCTGAAGGTTGCTGAGGGTGAAACAGCATGACCGATAATTTTT
...
122289 GGTGATTTCTTCTGATGATGACAGAGTGAAGCGCGCTCCAGAAAGTTCTGCGCATTTGCTG
122339
122349 AAGCTGATTCATTTTACAGCCGACACACAAATCATTTTTCGAAAGAAATAACAGAAATG
122409 CACGCGAGCAAAAGCCAGTAGATGAGTGGCTTTCGATGAACTGGAGCGTAAATCG
122438
122469 TTAACGGCGCTCTGTTGGCGGTTTTCCTTATTCGCTGAGATCGCAAGAGACAGCGCAAGC
122470
122499 TTAACGGCGCTCTGTTGGCGGTTTTCCTTATTCGCTGAGATCGCAAGAGACAGCGCAAGC
122529 GCAGCAACATGCTGTTGGCTATGCAATGCAAGTCTGCGAACCAGCAATGGAAGCGTACGCC
122538
122589 ATCAACCGCATGACTGAAAGCGAGGAATTTGCTTATTCGCCAAGCAATGACTGCAACG
122606
122649 CAGAAGTAGCAAGCTATTTCAGCGCATTTTTCAGCAACTGACAGACCATGCAAAAAACGGA
122669
122709 TCGGCTGGCGGCTTTCGCTCATTTGGCGAGGTCATGCAAGCTGAGGTTTGGCGCTTGGAG
122757
122769 AAGGATTTGACCGCTCAGCGGCAACAGAGAAATGAGCAACAGGATCCCATGCGTAC
122770
122829 AGATGCTGTCCAGCAAGGTCTGTGAAGGCTCTCTGTTTGTCAATTTGGCGTCCGCT
122877
122889 AAGTGGGGAACGACGCTATACAGCCAGATGCAATCAATTCGCGAGTGCATGAGAAA
122900
122903 AAGCCCGCTC
...
123000 AAACGTGAGACAGAGTCAAGTGTAAACCCAAATTTTTTACCTTCGCGGCAAAAT
123050
123060 GACCGCATGACGCTATACAGGGTGAATTACAGTGTACTTCAACAGGCGATGCAACA
123060
123120 GCTAATGCTTGAAGTAATCGACCTGTTTACATCGACGACAGCGCGGATTAATCTCTG
123124
123156 GCTCAATTCGTACAGCAAGCCGTCGAATCAAAAGAGAAAGGATGCTTGGCATGATT
123180
123186
123240 CTGTGCAATTACCTGACACTAATAGACCGCTGAGAAAGCCGATCGCAAGACCTTCTTAC
123285
123300 GGCATGATCACCAGAAAGACTGAAGAACCTTGCCAAAGACT
...
123350 TGTTCGCTTCTGACGACGCTTAACCGCGCACTGGAAGCCGAACCAATTAACGCCACT
123400
123410 ACCAAGTACTCAGAGATACAGGCGAGATGAACAGAGATTGCAAT
...
123485 TTTGATGACAGTGTTCACCTGCTGTAACCGAAMCTAATCCTTCGCTCAATTCGTCATG
123535
123545 CAATACCGGACAGGTGATTTGCAATTCAGGCAAAATGGCGCTTATTAAGACAGCAACA
123579
123605 GTCTGCTGAATAATGCGCCGACGTGAACGCCGA
...
123782 TTATGCGATGGCGGATTCCTGAACGAATTTGAGATTACGCCAGATGTAACGTTTAA
123832
123842 GGGTATGATTAATCATCATGTATPACGGCGTATTTCTCCATCTTCTGATGACGTT
123896
123902 GACATTTGTAATGAAGGAGATGAAGCAATGACCATCTACATCACT
...
123961 ACAGGCTGCTGTATTCGACAGGCTTTTATTTTGGGGAGAGGAAAGTGAACGATAGCT
124011
124021 ACCGACAGTTTGAAGAACTGTGTCAAAAGACAAAGCCAGTTACGCGGAGACGATGAT
124071
124081 TAAAGAGTTTGCTGGGTATATGCGAGGATCGCGCTGTATTTGAAGTGAATG
124158
124141 ACTGGCCGAATCGAATGACGACTTTTGGAAAGATGTTGAAGAAAGTGTATTGCGATG
124158
124201 GTTATGAGATG
...
124256 CGCAGAAATCAAGAAAAAATTTGATTAAGCAATATCATCTCAATTAATGCTTTTA
124306
124316 AGGCATCAATGCTCTGTAGGAGGTAAAGCGCTCAAGGATGCCAATTAAGCAGCA
124326
124376 GAATGA
...
123180 GCTCAATTCGTACAGCAAGCCGTCGAATCAAAAGAGAAAGGATGCTTGGCATGATT
```

118244 TACACGCTTGTATTTATAGTATGCTTTTAAAGCGGGGATTCCTGGCGATGT
118247
118304 AAGAGAGATCAGAAATTCGCTGTTAGTAGTGTATCTATTATTATTT
...
118395 AAAGAAAAACCGCATGAGGCGGGTTGATGATTAGCGGTGGCTTGGACGATTAATG
118445
118455 CTCACACACATGTTCAAGATTTCCATGCTTATCCGTTACATAGCA
...
118737 ACGTAAACCTTCAGAAATTGAAGAGATGCGTTGCGCTCTCTGCTTGTGATATCTGTTAGGA
118787
118797 ATGCTTTAATCGTTACTGTAAACCCGTGATTGAGGTGAGAACTGGAAGATCATAG
118806
118857 GTATG
...
118977 AACTTTTGATGTCACCTTTCATGCGGCTTTTATATGCAATAGCTGCTGCTGAC
119027
119037 TTCGGCTCAGTTGGCTTGAATGATGCTCGGCAAGCTCATTCACATCAAGCTGAGTAACA
119059
119097 TCACCCATTTCCTCTCT
...
119134 GAGCGAATGAGTTTTTAATTTGAGCGGTGCGCTCTCATTTATGTTGCGGACATAA
119184
119194 GCCTTCAAGGAAGTCCCTGATGCTGCAAGCTAAGCTTTGACTTGGCATTTGCTCTC
119228
119254 TCTTAAGTTATTGACTCAGATCGAGTTTA
...
119507 ATTTTCATCTGTCCAGAGAGCGGATGCACCTTGATTTCTGTTGATGGCAGACGCTTTT
119557
119567 CACCTGCTCAGCCTGCGCATACAGAGCGTCTGACCGATATCTATGCTCGGTCATTTGA
119591
119627 TGTATTAGACATATCCAGATCATTTACGATATATCTCAGA
...
119704 TGAACGCTGTGATCTCATGGCCATTAACTCATCTATCTCTTTACCCAAAGCTCTT
119754
119764 CAGGCAATGGCGCGCATTAATTTCCCAACAGGAACAACCTCATTTGATGGATGGATCA
119777
119824 TTGATATTT

...
119874 ATCAGTTTCTTGAAAGTAATTCATCCACCACCAAGTCTGGCTATGCAGAAATTCGCGGCG
119924
119934 TCAACAGCTTGTCTCAGGGTCAACCAAGAAATTAACATCCCGTCAGGAA
...
120571 TGTAAAGCAAGTTGAGATGCACACAGAGGAGCTATCGACATTTGGAAATCTGGCTGCTT
120621
120631 CCCAGACCTTAAACATGCGCGTCTGCTTGTGCTGCTCTTAACAAGTTA
...
120725 GCCGCTTAATTAAGCGGAGCCGCTCTTGTATACAACGAGCATTCGCTCCTAGCTGCTGAA
120775
120785 AGCGAGTCCCAAGATATCTGACCAACTAAGGCCATATGCGTTTCCA
...
120902 TCACAGCGAGAAATTTATCGCGCTGAAACGTGATTACTCATCAACCTGTGACCGCTTACC
120952
120962 CAGCGCGGCTTGGCAAGATGATTTGGCTGTGCTGATGAAATGA
...
121209 AAACAACATGATCCTTACAGAGACAAATTCACAACAACATACTAGGTGATCCCTTCT
121259
121269 GTCCAGCTTCACAGCAGCTGTGCTGATTCGGGCTGAGTTGAAAAA
...
121544 GGAGAAATTTCTGTAGGCAAGAGGAGACTACTTGGCGTAGAAACTCTGCGGCTTGA
121594
121604 ACAGACTCAAAACAGATCGGTTCACTACCTGCTTAGAAAAATTCA
...
121665 TTGACTGGGTTTCACACGTAATTTCTCAGTTTCTCGGTAGAGAAATATGACATTAATC
121715
121725 AGTCAAAATTTGTATCAGCAGAGATTACCAAGAAATTACCACCTCAAA
...
121916 TCCGTTAGCTGTCAAGATGTTGTGATGCTTACACAGAAATCTCTCTGAAGCGCAAAA
121966
121976 ATCCGCGCACTGAATGACAAAGCTAAAAACAGATCCGAACGTTCTGGCGCAAGCGCGA
122014
122036 GTGATTAACCCGCAAGCTTGAACGGCATGGGTTCA
...
122044

```
115850 TCCATGCCATCAAACTGCTGTTTTCATTGATGATGCGGGACCGCATCAACGCCACC 115900
115910 ACCGAGCATGCGCTTCTGCTTATCAGGGAAGCGCTAAATTTCTTCTGTCACGGATTA 115915
115970 AGCCCTACTGGTTGGCGAGATCAACAATGCGANGAAGTGGCGCATCGGATCAGCT 116014
115984
115987
116030 TTAATGCCCTGGGGAGAGAGTGCATCTCCTGTGGTGCAGAGATTCATGCCG 116035
116090 ACACGTTACAGCAGCTTCCCTGCCAGCTGTGAGTGTACTCATCCGCTTTATACCT 116114
116150 CTGAATCAATATCAACCTGG
...
116182 GTTTCACCAGTACCGGATGTTCTCTGCGATGCGCTCCTGAACCTCAACATCGTCA 116232
116242 AACGACGGGTAAAGGCTTTTTCGTGCGCCCGGCGGTGGCAAAATGATGATGAGACT 116276
116302 GATTCAAACAGGTGCTGGGGCAGACCTTTTCCCATGTCTCTGCGCAGTTCTGCTTTC 116337
116362 TCTTCACGGGCGATCTGCTGTGATGACGCGTCACAGCTCTGAGCCCTCAAGATCTCTGA 116411
116422 ATGTAATAAGCTTTCATGGCTGACACTCCTGAATAATGCTGTGAATAATCGCCGGGAAT 116430
116482 GCGGGGTGATTAAGAAAAACAGAAAGGGGTTAGTGAATGCTTTTGTGATCTCAGTT 116468
116542 TCATTAATTAATATCCATTTTATTAAGCGTGCAGCGGCTCAGCAAAACATCTTTTCATCGC 116567
116602 CAATAAAGTGGCGATAGTAATTTAGTCTGATAGCCATAAGTGTTTGATCCATTTT 116612
116662 GGGACTCCTGGCTGATTAAGTATGTGATTAAGCGCTTTCATCCATCCGTCAGATATTAGCG 116693
116722 GTGATTGCTCAAGTAAAGATTGGAA
...
116769 CCGTCATGGCATAATTGCATGTGTCTCTTATTATTACATAAGCAAAACGCTCTGA 116819
116829 GTGAAGCGTTATGTGTATCGGTAAAGCGCGGCTCAGCGGCTTGATAGTCAATCATC 116833
116852
116889 TGAATCAATATTCCTGAT
...
116968 CCATTTCCATCATTCAGTGCAGACTCAGACACAAACCATATGCAATTAAAGTCTTGAA 117018
117028 ATTGTATTAAGCAGACATGTTGCCGACATGATTATATACAGAT
...
117206 TGATGATTTTTTGTATGTTTGACATCTGTTCAATATCTCAGATATAAAATCGCCCTCA 117256
117266 CATTTGAGGGCAAGAAGATTTCATTAATCAGAACAGTGGCTCTCTGTTAGTTACGA 117325
117326 GCGACATGCTCCGTATTCACATCTGTGGAATGAATACAGATGCTGTGTTTAT 117330
...
117414 CAGGACGCTTGTGTAATGTTGACAGTATCATAGTAATTAATGATGAGCGCATTA 117464
117474 ATTATGAATTCATCGTTTGTGCGGTCCATCTCCATCTCTTGTCTATTTGCCATCTTGGG 117485
117534 TCAGTGTGTCAGCGGCGAAGCATAAACAGCCACGACCATGTTTAAAGAACT 117539
117594 G
...
117658 ATTCCATATATACCTCAAAATTAAGTGTGTTGCTGCTAATTTCAATTTCTGGCGCAAC 117708
117718 ACAAGTCAATCTGTCTGTCAGTTGTTGGATTACGTAACCTGCCG
117764 CATAAAGCTACATTTGGAAGACAAGTTGAGCTTCAATATTTCTGTCAGCAAGTGTCA 117814
117824 GTGTTATTAATCTGTCTCTCATTTGCTGTTGCGTTTGCATTTGAAGCACCTCGTATG 117913
117884 GGGTTGGCCTGTGTAGTTTGTGAGAGCTAATGCGCTCCGACTTTCAGAGTTTGGCGGAC 117913
117944 GAGCTTACGGGAGAACGCTGGGTCCTTTAAATTTGTTTCTGGAACATATATTCCT
118004 CCCGAATTAACCTTTGGCGATGCAATCTGAAGCTCTCTGAGAGGTTGCTGGGCAAT 118018
118064 GCATCCACAGCTTATGTGTTGGGTGATCTGGCTTTTCAGCACAGTATGAGAGTCA 118122
118124 CGTTGTTAAAGAGCCTGCCAGTCTGTTCATTTGGCTTCCAGGCTCCTGCTGATGTTA 118127
118165
118184 AATAGTACATATGACTTCACTTGTGTCATATCAATTTGTTCTAAACGGGGCTTTT 118234
118234
```

```
112481 CCTCTTTTGCACAGCCAGCCCATGACGTACCGAATATACTCACCGAAGTGTGGCGACTTAT |-----|
112531
112541 ATTGGCATCTGTACTGAGGCAATCACAATAGATGCATGCT |-----|
...
112694 TAAATATTTCAACACCCCATCCGTGGTGTGATCTCACCCACTGCCAGACGCCAGGA |-----|
112744 112753
112754 GCGTTTCTAGCACTAACAGAAATCTTGTCTTTACGGTTTTCATCGCTTTTCTCTCTGC |-----|
112758 112796
112814 ATCTCTTTGCTGCTGCTGCTGCCCTGCAATACGGTATGGCGGCTTGGTGTGCGGA |-----|
112836
112874 TGATGTCTTTGCATTTAGCGCAGAGCCCAAAAAACA |-----|
...
112969 GAATGCCAGGCGCGCTCGGCTTCACCTTGTCCCAAAACCACTGTGAAGCGGCATTA |-----|
113019
113029 GCTTTTGTCAATCGGTGCATATTTGGCATTTAAGTAGCCCTGAGC |-----|
...
113097 AATTGCCAGGCGCAACAGTGTGTAAATATCCATGCTAGTCCGCTTTAGCGTCCGA |-----|
113147 113150
113157 GTTGCCCTGCATTCGATTAATGGCAGCCGTGCATGACAGAAATGTTTCC |-----|
...
113245 ATGCATTCTGCGCTTGTAAATTTTGGCTTGGCGCTTCTGAGGCTTCAAGCAACGCTGCT |-----|
113295
113305 TATCGCCTAGAGCTTCTTCAGTTGAGCAACATGGCATTCATATC |-----|
...
113381 AACAGCAGCTTCCCAACCGTAAGCGCAATCTCCTGTTCTCGTGGCGCGCTTGAT |-----|
113431
113441 GTAATGCTGTCTTCTTTTCGTTCATCAGCAGTGCAGCAGCGGTAGCGGTTAGCCTC |-----|
113501 TGGTATGAATTCAGCGCTTGGCATTAAGCCTGATCATCTGATTCATTCAGGCAAGTTAACATG |-----|
113513
113561 ACATTCGG |-----|
...
113749 CTCACGATCCTCCTGGCACTTGCAACAAGTCCGAAACCTGAAGGCCAGCGCTCTTCG |-----|
113799
113809 CTCATCTATGGGATCGGCACACTCACAACAATGAGTGGCAGATATAGCTGTGTGCTCAG |-----|
113821
113869 GCGGCGCA |-----|
```

```
...
114004 AGCCTGTAGCTCCCTGATGATTTTGTCTTTTCATGTTCAOCCTTCTTAAAGCGCGTTT |-----|
114054
114064 AACATGCCGATCGCCAGACTTAATATAGTGTGTGAATCCCATTAAGGCTTACCGTTTCG |-----|
114073 114109
114124 CGGTCTTCTTCATGATACCTACGGCAAAATGTCATGTAGAGTTTTCGGAAGACTGTCT |-----|
114159
114184 CTGCTTTTGTGATTTTCAGAAATTAAGCTGACGGGCAATGTCTGCAAGGCGCTTTCCTGC |-----|
114231
114244 TGAGTGTCAATTAACAAAGTCCCATGTGCGCAAGCATTAAGCAC |-----|
...
114626 AGTAGACGAAGTGAATCGCACCTACCGGAATATTAATCGTAGAGATGCTTCATCGCATTTG |-----|
114676
114686 CTCCCAAAATCAAAACCAATTTTCAGCCAGTGCCCTGTCATTTTT |-----|
...
115041 ATCGGGGATTCATTAACATTCACGCGCGGAAGTGAACTAACAGGCTTCTGCGCTTTC |-----|
115091 115094
115101 TCGTACTGTTTTCCCAAGCCAGTGTCTTAAAGGTTAATCTCCGAGCCACACCGGTGCA |-----|
115130
115161 ACTCGGCAGCAGGGTGTGAGT |-----|
...
115212 TTTCTGAGCGGGCTTTGCTATCATCAGTTTGAACTTCTGAAGCGGTGATGACCGGAGC |-----|
115262
115272 CGTAATTTGTCCATGCATCATCCCTTGTTCGACAGCTTCACGT |-----|
...
115380 TCTGTTTCAGGAATCAAGAATTTCACTGCTTGGCCTGTGTGATCTGACGATGCGC |-----|
115430
115440 GAATGTGCGGCGAAATATCTGGAGACAGCGGCATTAAGTCATCCATGTTTAT |-----|
115480
115500 CCAGGGGATTCAGCAGAGTGTAAATCTCCTCATAGTTTCATGCAACCGGAGTGATGT |-----|
115504 115540
115560 CGCGTTCGCGCTGCGCTTCTGACGTATGACGATATTTTCGACAATGCGCTCGGCTCAT |-----|
115572
115620 CCTTGTC |-----|
...
```

109228 AACAAGTGCACCTGTCGCAAAATGTTCTTCCTACTCGCGTAATTGACAGTGAATCT
109288 GATGCTGCCCTGCTAAGACGAGTGAAGCATCGCCTGGTTTAAAGTCTTCATTTGCTGTG
109348 CCAGAAGATGTATAACCCGGTT
109762 CGGTGATAAATCTTCTCGATTAACAGCAGATCTTCTTTTGTAAATCGATGACGCCAGCC
109822 TGCCAGACCAAGCCTCGGTAGCATATTTTTCATCATGATTAAGCATTTAAGCGGT
109882 TACTACCGATTTCCATTGCTTTACAGATTAAGCCATTTCTGTGCCAGTTTCA
109957 GCCCAATAGCTTACGTTAGGTAGTCAGGGAATCGATCGGCATATATCAACGTTCCGTG
110017 TTCTGATTCGAAGCAGACCGAAGCCCTTAATCTACAGCAGCATCTTCTGTTTTCGCTG
110077 TATCGGCGACCGAAGATTTTCGGCACCTCGTTTACCCCTGTACTTAACGTTAATTCGGA
110137 GTTTTCTCCGTGATTTTCACAGCCCGGCTGGATAGGCTGAATTCGCCATTTGTTCTCTCTG
110197 CGTCCAGAGCAGCAGC
110273 TCCAAGTGTGATGTAGGAAGATGATGAGACATTCGCTGTTTCTTGTGCGTTCCGT
110333 CGGGGGGATTTGTTTATATTCTCTGCCGTTCAATCCAGCAGCTGTCCCTGGCGGTAATA
110826 TTCCGCCACACAGTGAAGATTAATCTTCAGGCTCTGGCTGTAATTTCTGTATACCTTGAAG
110886 GGTGAATGTGAACAGCGCATTAATCATCTCCATAAACAACACTC
111049 ATGTCTTGTGATTTCATATCCGTATAGGAACATCAATTTCTTGGCTGGAATTCGTCGCGC
111109 TTGGATATCCCTTGTGATTAATCAGGCGGTATATCTCGCGGAATATATTCGCTTATTC
111169 CAGTAGCATTAACAGCGCATCTTCCGTTTCTCTCCGGAATCATGCAATCGAAG
111185

111229 TATTACCTTTA
111285 CTGTAGCGGGCATCCAGTGAATGCTTGTCTCATATACATTAACCGGCTTAATCGTTGAT
111345 CTCCGCCCGAAGGTGCTTCCGGTATAGCGTGGCGAGCATATAG
111483 GAGTACCGGAGAGTTTACCAGCCCTCATAGCGGATTTTCATCCAGTGCCTAAGCGCTTCA
111543 TGCTTACACATCCGCAATCAACGCTTAATTTTCTTTTGTCTGACAAACCATTTCTGGA
111603 ATGACAGCTTGGCAG
111745 AGTGGCAATGTTGACGTGTCATAGTCAGTCAACGCAATTTTCATTAAGCAGCGGCTCAAT
111805 ATTATCTCGCAGCTGCAGGCTGCCGATTCGCTTTGATTTCTTATATCATTTTCTTGTG
111865 GGTGACGTTGTGATGTTAGTGTATGCTCTTATCCACTTCAAGCGCATCTTTCAGCGTGAATGCC
111925 GTGCCAATATCAGGCCAAGTGTTAACCCAGGCGCAATCAATGCTAGGCATATTAAGCGT
111985 TGCAGTGTGTAGGCCCTTATGCTTATCAATGCTGGCAAGTGTGCGGCTGTATAGCGCT
112045 CAGCGTCTTAGA
112060 ACTATAAATCTTCTCGCGCTGCCCTGAATTCCTTTACAGCCTTCACACAGCATTTCCA
112120 CGCTTTTCTGTCTGCGCTCAGGTGATTAATTCCTGCAAGTGTGCCATTTCAGTTT
112180 CCTATATAGGTTAATTTTATTTGACAGTGTGTGAACGAGCCCATTAACAGCTGTAT
112294 GGGATATCGTGTATCATACAGCATTAACAGAAATGAGAAACAAATTTCTGCTCTAC
112354 TTTTTCGGGATTAATCTTCTGATTAATGAGGACATCATCTCTC
112405

```
106334 CAGATGAGATCTTGATGATGTTATTTAATAAAGCTGGCAATCATTTATGACGCTTACC |-----|
106394 TCCCTCCTGAACCTTTTGGCTCAGGCTTAGAGTTAACAATTAAGGA |-----|
...
106509 TGTATTAGTCTCACACTTTTATTATGTAATATTCCTCTGTTATTTGTTACGCGCGG |-----|
106569 ATATTCTGTTAGCCACAGGTGCAATCATCAGCGCGGTACGCGAGG |-----|
...
106731 TGTACACTCTTTAGTATGATGATTAACAAGAAGGTGCCAGAAATGACAAACGCTTATC |-----|
106791 GCGGAGCTGTGGCAGGGGTAGTTTACTCGCTTCAACGCTCAGGCAACAACGTTACCG |-----|
106793 |-----|
106851 GAAAGCTATCAGCTACAGCAAGTGTCTCATGAT |-----|
...
106890 CATAACTTACGTGGCGCGCTGGCGAACAATGGCAGTGTGTGGACAGTGCAGCGCGAAT |-----|
106950 AAATGGCCCGAATGGAGAGTCCCGGGTGGGCACTCACCACCAAGTGGCGTACTGGA |-----|
107010 GTGTATATAGG |-----|
...
107281 AAGAGCTCAGCAAACTCCAGCTTACCGACAGCTACAGCTACTGAAAAAATCGTTAACT |-----|
107341 ATAAAGATTCCTCCCTGCTGTAAAGAGAAACAAGTGTTCGCTGTGATGGCAAAATA |-----|
107401 CCTTAGCGCCAAAGTATCAACAGAACCGAGTGTTCGCGCGCGCTGAAGTCCGCAACT |-----|
107461 C |-----|
...
107682 AAAATCAAGGTGTAGTTGGCATGACTCCAACATTCCTCTCTGTTAAGCGCGCTGAT |-----|
107742 TTCAACCGTATCAGTTGCATGACCAAGAACGACGACGCGATTGGCGCAAAATCGTT |-----|
107802 TTCCAGCGTTCGCTGACAGCAAAAGCAATCGCATTTGATGAAAAATTGAATGTGTAC |-----|
...
107874 CAGTTACGTAATGCCGATGCGTTAACCTTCGAGGACCTGCGCAGCGCGCTGGAA |-----|
107924 |-----|

107934 TTAAGCGGTCCCGATGAGAGCTGATGTTTCTGCCCCAGATGATTAAGTTGATAGCGTG |-----|
107952 |-----|
107994 TTGAATGAAGCGGTGAATAATACAGAAAACTCCCCGTTGCTACGGGGAGATTCGCTGTT |-----|
108054 |-----|
108114 CAAAGGCTTGTACTAACACTTCATCGCTACCTCTTCCCAATCTTTTCCGCATTTT |-----|
108174 CGTCGTACCCGCGACCTTCGAATAATGGCTTGCGCTATTTCCGGGACGATTTGCGCAGAG |-----|
108234 ATGCCAATTCACCGACGTGATGAGCTTTCGCTCTTG |-----|
...
108394 CCGCTGCGATGTGTTGCGCGCTAGCGGCTACCGCGCGAACCCTGTGAACGTCAAAT |-----|
108454 AATTCGTGCGCTGCGTAGCCAAATGGGAGCAATCACCATGCCGTATGGCAAGGTCGTC |-----|
108514 CAGGTGATGTGATGTTTGTTCCTGACCGCGCGCAGTACGGGTGAACATAACAGCTC |-----|
108574 GCCAGTTTCCGTATAGTGCGCCGGAAGCCACAGCGCGCGCTGTGCT |-----|
...
108634 CGCATTTGACCGGACATGTTGCAAAAGCGGAGTGTACCAAAAATATATGCGTCTAA |-----|
108694 TTGGCGAATCTTTCGCGGGGTTCGAACCGGTGCAGTTGCGTTTACCGCGCTTTTCA |-----|
108754 AATAATTGCGGCGACATGTTTCCGTTACACGCTTAACGACACTTCAGCGCATCCACT |-----|
108814 TTGCTTGACCCCTGAGGAGTCCGCGTGCATTCGTTCAACATGTCTAAGATATGTTT |-----|
108874 ATCATTTGAAAAATATATTTTATTTATTTATTT |-----|
...
108972 GTACAGCTGCACCTCAGGCATCCATGACACATTAACATTTAGACCATGTCGCTGGAG |-----|
109032 TGCCCATCTGTGATGAATAATGTGGGTTTCTCCGGAAGATA |-----|
...
109108 GATACGCTTTACGGATCGATACCGGCTTTTATTATTGCTGAATCCCATGTCGCTCGA |-----|
109168 TGAGCTTACCGCTAGTTAATACCGCGCTTGTGATTTCTTGCGAACGATTTGCGACAGA |-----|
109211 |-----|
```

```
103558 TAGTTGAGGGGACCAACAAACAGATPAACGTCGTTGTTAGCCCTGATTAAGACGGCTTTA |-----|
103566 103585 103588 103617
103618 GCGTCGATCAAGGCAATTAAGGACCGCGCGGATGCGCGTGAACACGGCTTAT -----|
103622 103625
...
103769 GCTGGCGCAGCAGGCGGTTTCTGCTTCAGGTGGCAATATCATCCATTAAAGCTACCG |-----|
103819 103825
103829 CCACCGGATCCCGGCCAGTCAGAGCCAGTTCATGACGAGCGCACCGCGCTTGCA -|
103830
103889 CCACGATGCGCGCAATGGTCGTCGAATACCAAGTTGTTTCTGATCTCAGCGGTTCAA |-----|
103905
103949 CCACCCCCCAAC -----|
...
104029 ACCGTAACTTAGCATTTATGCTTCCCAATCTTTAGTGTGATCAAAAGACAGCTGGG |-----|
104079 104087
104089 CGTCTGCCAGTTGCTGCCACAGCGCGAGTGTTCATCCGGTTGCGCGGCATCAACA ---|
104092
104149 TTTTCAGTACCGCATACAGATCGCGGTGTTTTCCTCACCAGACCTTTGCCTTAA |-----|
104168
104209 CGGCGATGCGTTGGCCGCGCTGGCTGCCAGCGGAGATAGTACAAAATGCTTTCTTCA |-----|
104215
104269 GTGTTGGAACGGTACTTTAGCACCCAGCGCGCTTCCACAGGGCTAAACGGGCACACAA |-----|
104295
104329 TTTCAGATCCTGGCCGACATATCAAAACAGCGGATGGGCAATATGAAATCACAGCC |-----|
104389 ACAATATGCGCATTTGACCGCGGTTTTCGCCCGCGTCCCTTGCCCTTTCAGCGGATGC |-----|
104393 104421
104449 GTTGACCATTTGCGACGCGTGGGATCTTCACATTCACGCTTTGGGATTTCTGTGT |-----|
104462 104487
104509 CGATCATGCCAAGGCGTATTAACCGGCAAGTTATAGTGTAGCTTATGCTACG |-----|
104522
104569 TAAGCGTTTCTTGAGAGAAATACCGCACTTCGATTTCAATATCGTGGCGCGGTCGGCGG |-----|
104571
104629 GCGGTTACGCGCTCTGGCGGCAATGCTGACCAAAATTTGACGAGAAAGATATCGTCAAAAT |-----|
104678 104678

104689 CTTCGCGCTTAAACTCTGACCGTGCATGCTGGAAGTGAAGCGGCAAAATTCGGATCGT |-----|
104693 104711 104744
104749 TCGGATGTTGGACATCTGATCATCTACGACCGCGGCTGTGTTGATCACTTAACACTTCCC |-----|
104749 104782
104809 AGGCTTCAGCGACCTCTTTGAAGCGGCTTGGCATCCGGTCTTTCCTGACATCAGGAT |-----|
104816
104869 GGTATTTCCGGGCAAGTCGACGATAGCGGCTTGATTTGCTCTGAGATCGTCCGCGGT |-----|
104887 104915
104929 TCACGCCCATGATGGCGTATATATCTTTAATTCATAGCTTATCTCGCTAAATCAAC |-----|
104966
104989 ACAATTTGAAGGAMCCCCGTAAAGSTAACCTCT ...
105045 ATTTCAATATGCCAACACAGAAATATGTTATTTGAATCATCGGAGAGAGGCTCCCTCAA |-----|
105095
105105 GGTGGGCGGCTGAACATATTTTAAACAGTGAAGGAGGATGAGATTTGATAGTTGA |-----|
105110 105146
105165 AACAGCACTTAAATTTACTAGTGGCTATTTATTTATTAATTTATAGCGCTATAT |-----|
105215
105225 ATGGTAATATATCATGGAATTAGTTGCACTGCAATATATTTTGAAGCAAGCCTGGAC |-----|
105282
105285 GATATATAAAATGAGTACGAAAAAATTAACATATTCATCACGCTTTTGAAGAACACA ---|
105287
105345 CTGGTGTGTTTATGCGGTGTATTAAGAAATCAGAGCTCAACAAGCTGAAGACTACATT |-----|
105379
105405 CCATGCTGGAATCGATTCTACTAGAAAA ...
105796 TTCAGAGTGAATTAATAAACTTATGATTAATAATACGCTCCAGTAGAGCGCATGCT |-----|
105846
105856 ATTGTCTGTTAAGGCGCATCTGTTTTTAAAGAAACATGAGTATTCATGAGGCTGGGCA |-----|
105905
105916 GCCATAAGACGCAAAAGCAAGTATACGAGAGTTTATTAATAATTTGATTAAMCAACGA |-----|
105971
105976 TCGAGGCTCTTGTGATAGCGCATTTATATCTCAGGGAAGATGTAAGACGTTATTTG |-----|
105976 106025 106035
106036 GCGTTGATCTACAATCTGACAACAAAATTCAGGGGAGGGTAGTCTTATGTATG |-----|
106040
```

```
100895 CGAGTGCCTTGTGACGCCAGTGTGGATTGSCATCGACGGGGATGTTCCATACGCTTCGG |-----|
100945
100955 GGAATGCTGGCGAAGAGCTATTGCTTGCAATGATAGCGCTTGACGGCGGAGATTACT |-----|
100990
101015 CTACTGTGCTGGCGAGTGATCTGCGCGG
...
101165 GTGCCCCGATCAGATGTTTATGAATATTACGGGAGTTGAAGGGAAGTGGCCCCGG |-----|
101215
101225 GTGAATTTAGTGATTAAGCATGATCCGGTTGTTACATCCACCA
...
101290 CATGTAGACACATTTGCGGTTAACCCGCAAACTGACGTGCCGCTGCAACTGGCGCTGGCG |-----|
101340
101350 CATACGTTGTACAGTGAACCACTGTACGACAAAACCTTCCTCGCTA
...
101569 CAGCGTATGCAGCAGCAGTGACAGTGGGCGTGATGATGTGTGCTGGCGGAGATCTG |-----|
101619
101629 GGGCAATTTGGCCTGCCAGTGTGGCTTTGGTTTGGCTGGCAGTATTAAGGCGCAGGC
101689 ACGCCGGGGCGTTAAGGCGTTTATTTGAGTGTTCCTCGGCTCTACGTGATTCGCGCT |-----|
101704
101749 GTTCACGACAA
...
101859 AAAACTGCGCGCGCTGAATAATGTGATTTTTGGCCGAGTAAACCCCTCCATGCGCATCA |-----|
101909
101919 GCAGATCAACCGCATTTTGAAGGCTGGGCAAACTGGAACGGTTATCGCCATAGATAA |-----|
101965
101979 CCAGTGGACCTCAACCTGCGCGCTTTGCCGATATGTCGTGCTGGCAGCCAGTTTGA
102039 GCGTAACGATCTTCGACCACTAGCGCAACCACTCCAAACCGTGCAATTCGCCATGAACA |-----|
102043 102085
102099 GGTTCGCGCGCGCGAGTTGAGGCGCGCAACGACTTTGATATTTTCGCGAGCTGTGCGG |-----|
102127
102159 CCGCTTTAATCGGAGAGAGCCTT
...
102206 GGTGCTGAAGCATCTGCGAGAGAGTGTTCAGCAAGGCAAGAGACGGCGGCTTCATT |-----|
102256
102266 TGCCAGCGTTTGTAGTACTTCTGGAATTAACAAGAGTACGTGAGTTGACCATCCGAGA |-----|
102270
102365 TCGAACCGCTGGGCACGCCAGTGGCCTGATTGAGATCTATTGAAAACCATGCGGATA |-----|
102415
102425 TGAACGTACGAGATTTGTCAAGGGGCAATCCGATGTGTGTGAGAAAATGCAAGCTCCACG |-----|
102433 102473
102485 GTGGGCCCGGCTCGCAGACGTATCCGTTGCATCTGCATCTGTG
...
102626 AGGATGCCAGCGCGCGGATTTGTTACGCGTATGTGATCGGCTTTTAAGCCTGCGG |-----|
102676
102686 GTCAGGTGTTGGCAGGGGAGTAGTATTTCTGACCGCTATGCACCCGG
...
102805 GGTAAACCCCAACGTTGTGACCAATTGACATGCTACTTCGCAAGCTGGCGGACGACAGT |-----|
102855
102865 GCGCACACTACGCTGGTGGAAATTGAGAAAGTGCACAGGAACAGTGGAGCAGGTAAGGCGG |-----|
102921
102925 TTTAACGGCCCCGTGAGATGTGGCGCAGTGGCAATATGTTCGCGGTGCGC -|
102926
...
103009 AGATAGCCGTGTGTTTACGCCCTGGCTGGCGAGTTGTTCTCCCGTGAACCTGGAGATGAC |-----|
103059
103069 AACTGACGCAATATCGCCAGTGGCGCAGATGGCTGAATGTTTGTGCTGTAAGAAAGCAAC |-----|
103080
103129 CGCGCGCT
...
103138 CTGCGCGGATGATGAGTGAAGTGAAGAAACCGTGTGCCACTGTGACAGTACGTGACGATGCCC |-----|
103188
103198 GTCTGGAACGTGGCGGCTGACTTTTGGCGGCTGTTTCTGATGACTGACAAACAAGCGCGC |-----|
103253
103258 TGCCGTATGCAATGCGGCTACAAACAGAGCAGCAAGATTAACGCTTTGTTAGTGAAGG |-----|
103258 103300
103318 CAGGATGGAACCAAGCGGCAATTTCAACGAAACCGACAGATCAATCTGGCATCTATCTCG |-----|
103364
103378 AGTTGTTGAGTCATCTGCAATTTTTCGCTGGGAGAGGGGACGTTCTCGCGCAAGAAATCG
103438 ACAGTTTAAGGCAAAAAACACTCAGGCGGTTGGGCAATGTTACCAAGATTGTTGGCGC |-----|
103463
103498 GTTGCCATCAGTATGACAGCTTTGTTTTCAGCGGCACTTAAGCAGTTATGTGCTGTGT |-----|
```



```
98402 TATCCGCGCTGCTGGCTGGCGTGGGGGATTTGCTCCGCTATTTTATACGACACATC |-----|
98462 AGCGGCTAAGA 98417
...
98477 ACCTTGGCATTTGCTGCTCACCAGTAATTCGCGCTGAGTTTATGCGGTAACGAGAAAT |-----|
98537 TAACACATGACGAGTGGCAGAGTCAAGGTTTTCACCCGACGCGAGAGAGCATACGACG 98527
98597 CAGAGCTTCACGGCTGAGAAATTTGCCCCGGATTGCTCACAAATGCCACCAACATTTCA |-----|
98599
...
98684 ACCATCCGCTCCAGCGTATGCGCGACACATTCAGGCAATAACCCGCAAGCGATAGCA |-----|
98744 GTTGCTGAGTGTACGGTTGAGCTTGTGCGCGAGGTGATTCGC 98734
...
98799 TTTTAAACCGTACACAGTTCGCGCAGATTCCACGCGTTTGTGACGTAATCGCTGCG |-----|
98859 CCCATTTCCAGCCCAACAATACGTTCAATCCGATGCTGCGTCCGTAACGAAATAATC |-----|
98919 CCCACCGTTGAGCGTTCTCGCAGAGGCGGGGTTAATCATCAGGCCATTTTCATCGGCAAG 98891
98979 TTGATA 98929
...
99186 CGTAAGCCGTTCACTGACGACAGTCCGCTGTCAATTTACTCATTTAAGATCGCTCAC |-----|
99246 TAAACCATATTTTACAGGGGCTATTTATCGGAAACTCTGAACGCGCTACGCGACCCA 99236
99306 GTGCTCGTTGGTCCGTACTGGCGGTGCTGCTATTTGGGATTGTGATTGCAATGCGCTGA |-----|
99366 TTGTATTGCCACACGTTGGGAT 99325 99332
...
99426 ACAGTATGCACACAGTGTATGAGAATATTAACAGATCGGTGCATTTCCAGACGCGTCCG |-----|
99486 GCGTGCAGCTGAATGCCATGACTGTATATCCGCGGATATTCAGGCAATGCTGAAGC 99476
99546 GCAAACTGAGAGCAGCAAGATATCTACAGACCTTTATTTGCCACATCATTTATACAC |-----|
99563
99606 CTGAAAAATTTGA
...
99634 ACTTCCGACGCTGAATGGGCGGATGAAGAAAAACAACCTCGGCAACCTGCTCTG |-----|
99694 CCATACTACATGCGATGATCATGCGAAGCAGCATCCTGAAGCG 99684
...
99876 GTGTGACACGCTGTACTCTATTTGATATTAAGCAATTTATGCGCGGAAGGCGGATTAAG |-----|
99936 AAGCTCTGTCTCTGCTGCTCCCTCGGAAGTAAACCTCTTAACGTGACGCGACT 99926
99996 GGCTGCAATTTGAATTTACCGGCTGAGACGAAAGCGCGGAGCTGACTACTACCC |-----|
100056 AATCCCAAGTAAACGATCTTTGTGTGCTCGATTCGTGTGATGTGACAGCAGATTA 100028 100035
100116 AAGCCTGAGAAAAACACCGTTGCCGACACCAATACGAGTGAAGCAATTGACAGCCA |-----|
100117
100176 CCGCGTGAATGAAGAGGTGACATGTGAACGATATCAACCCGATCTGGGCTTATGGCG |-----|
100236 ATTGTTGTACACGCGACCTGTGTAACGATGCGCACAGGCGGCAACCGGAAATCTCTCACTTGG |-----|
100296 ACGCTAAGCGCTGATTTGGCACGCTCAACGGCATGATCGCTTTACAGCCTCGA 100205
...
100371 TGTGAATATCTGCAATGAATGCGTCTGACACCGCAGTAAGGCTACGCGCATTAAG |-----|
100431 AGAAGAAAAAATATGAAACAATTAACGATCTCTTTCAGGCATCAGCTGGCGTTCTGCG |-----|
100491 ACAACTCGCGGCTTAACCGTTCGCGGGGATGCTGGGCGGTCAATTTGTTAACGTGCGCGG |-----|
100551 TGGCACTGCGCGCAAGCGGCACTGAGGCTGTGATCTCGAAGAGGGCAATCTTAACCGG |-----|
100611 GTCGCACTGGGGGCTATCCGCGGACGCGTAAGAGATGCTGCTTTGTCGGCGCAAAAC |-----|
100671 GTTCGAACGTGATAATATATCCGTGGAATGATTT 100649
...
100728 CAAGCGGCGGTATTGCTTATCCGATGTACGCGGTGACTGGCTGCTAAGCGCCATCT |-----|
100788 GAGCGACACCTCCAGCGCGGTATTAACCGTTTGTGCGTGTGAGTGGATGAGC 100778
100789
...
```

95928 TCACGCTGGCGCTTAAAGTGGCGGGTATCTCCAGTAGCTGTGCGGTGCCGAGATAACCG
95936

95988 TACAGCGGTGTGGCGATCTCATGGCTCATGTGCCGCCAGAAAGCCCATTTTGGCTGGCTG
96015 96027

96048 GCTTTTCTGCTTCCGCCCTGCTGTGGGTGTTTC
...

96095 CTGCATTTAGCTGTGTACGCGCTTTGACCTGCGCCGCGACGCTGTTTCGCGGTGGCGATTACAG
96155 TGCATGAACACATGCTGCGAAGACCATTCATCAGCCGCCGATGCTG
...

96217 CAGGCGTTTCCGGGAAAGGGAGTCAATATCACCGTCCAGACAGCCGTTGCAGCGCCTGGCG
96277 TTTGTTGGCAAGTGTGACGCGCTGTGACTGTGCTCTAAACGTGGCGCAACTGGGGATGTAA
96337 GTGCGCAAAAGTGAACCATCCCC
...

96432 TCAATGGTGTGACACGACGCTGACTGTGCTGCTAAACGTGGCGCAACTGGGGATGTAA
96492 TTTTGTGCGAGAGTTTGTAGTGATTGCTGATTTCACTGTCTGTGTAACAGCGCCAGCC
96552 AAATCGCTATATTGGCTAACGGTAGTTAACGTTGTTGGCGACCTGAGCACGAACCTCCCGGA
96534 96579

96612 TCTTCAATGGGTATTGGCGAGGTGCAGAGATTTTTCACCGCAATTATTGAGCTGCTTTCC
96672 |-----| |-----|
AGCGTTGGCGCATTTTCTGGATCTGCTCCAGCCCGACGATTCATCACCAATTTGCTCCACC
96672
CGCAGAGCGCTAAGGCGCAGTTATTCATCTGTGTTAACTACTCAAGATGATATCATTC
96732 |-----| |-----|
AGCGATCGAGTGCACATTTTCAGCAGCGCTGAGCTGATCTTGTTCATCAATCGTAATATC
96852 |-----| |-----|
CCGGCTGGGTGCTCCGGCGAGGTGGCGCATTTATTCGCCCTGACCTTGCCCGACGGCGT
96912 GCGATCTCATCGGCGGCGACGAGCTATCTGCTAGCTAGTGCCTGTTGTTGCTGGCGCAGT
96972 |-----| |-----|
TGCAGAGCGCCCGCCACACAGATTCCTGCTGTGACGTAAACGAACGGAGATCTCTGCTCC
96976

97032 TGTGTTTCAATACGGTGTGATCAAAACCTGTTCCCTTAACGCTTGACCAACGACATTA
|-----| |-----|

| | | |
|-------|--|-------|
| 97092 | ATCTTCAGGCTTTGTGGCGTAGAGATTTCCGCCCTG | 97071 |
| ... | | |
| 97137 | ATCTTTTCGTTATTCGCACTGTCAGATTTCTGCGGCGAAGATTCCACAGCGCTGCT | 97187 |
| 97197 | TCGCTCAACTGCGCGCCCATATTTCATGTGTAGAAATCAATGCTGAG | |
| ... | | |
| 97264 | TGATTAAGCCGAGTTGTACCATCCACACAGGTTACTGGTCAGGGTTTAACAGCGCATCA | 97314 |
| 97324 | GGGCAAGCCCATCCAGAGTCTTCGGGTCAGGGTTAAATTACAGGT | |
| ... | | |
| 97578 | GTATGAAGAACTGTTCGCGCTTTATCCACGCTGAAGATTCAATTGTATTCGTTGAAC | 97628 |
| 97638 | TATGTTATGCAGAGAGCTGCTCCGCCCTACGGTGTGGATTTAAAG | |
| ... | | |
| 97734 | GACCAGTGTAAACAGTGGGGCGCAGAGGCCATTTTGCTCGGTAGTACACAGAGTCATTT | 97784 |
| 97794 | CCCGATCTGCAAAAGCAAGTAGCAAAATCTGCGGCTGATCGAGCTGG | |
| ... | | |
| 97862 | GGTGAAGAACCGCGGTTGGTGTGCCCTGTGTTTCAGATGGGCTATCAACCTGGGCGATATCT | 97912 |
| 97922 | GGTCAATGGGCGCACGGTAAACCACTGATGTGCTGTGATGCCCGGACCCGATTAAGCG | 97977 |
| 97982 | CGGGGCGCAGTAAGAGATGTGCGAGGGTTTTCGCGCAGCCATTGCCGGAAGCCCGGTGCG | |
| 98042 | TATTGTGATATTGGCGCTGTGTGATTAACCATATTGAATAATCCAGCGTAACTGTTCAGAGA | |
| 98102 | GATGCTGGAAAGCGCATCCAGAAATCGAGGTCGTTGCCGGAACGGCCATTGGCGGAGAGGC | |
| 98162 | GGCAATGGGGGAAGGCGTAACCTGAAGAGCGCGCTTACCTGTGTGTGCTTTTATCTTTC | |
| 98222 | ACATCAGGTGTATCGCGGCGTGAAGCGGGGAAGAGTATTATGCTGCCAGCGATCAAT | |
| 98282 | GGTCTGGCAGGGGAAGTGGCAATTGAGCAGGCCATCAGGCAATTACAGGGGCAATCGGT | |
| 98342 | GCTGTGATTAATGCAGCCACCGATTTTATGTTCTGAGCGCGAATAAATGCCGACCGGAAACA | |

```
93244 AAGAGTTAAGAGTATATTGAAGAGTTAAAAAACAAGCCGAACGGAAATCGCCGACT |-----|
...
93304 TTATCGCGCAAAAATTCGCCGAGCTGAACAAGAAATACAGGAAGAAGACTCTGTGAATTC |-----|
93318
93364 GCTTCACCGC
...
93443 CCCCCTGGAGGTGATGTTCCGCTTTGTTGTAATTTAGTGACCAATCATTTGGCGCTTGCG |-----|
93493
93503 GAATTAAGCGCTGCGTACAAATTCCTCCGGCACCGGCGCTTTGCCATACCTCTGCATACATTG |-----|
93509
93563 CGTAACCAATACCGTAACAATAATCCCAAGAACAGTAGCGTCATTAACCAAGCCAGACA |-----|
93601
93623 ACGCAAGGCTTTTATTATTTGCCGAGGTTTTG
...
93779 CGCATTTGCCAATCGATACAACTTTTCGGCATTCGCGAATCTTAAACGGGCAATATA |-----|
93829
93839 GCCAAACCAAGCCCATCAGCGCGCCATATGGGCAAGATATACGACACCAGGCATGGCGAA |-----|
93878
93899 TAAACAGGCTGGCAATCAGCAAAACGGTCAAGCTGATTAAATGTGCGGTCCCATATGAC
93959 GAAAGAAATCGAGCATTTTAAAGTCCATCAACGACGCTGTAGGGGCAACATAAATAAGT |-----|
93978
94019 GAATCGCCTGAGCAAGCATCATATACGCGATATAGAGAAAAAACTAACAGCAATAA |-----|
94021
...
94099 TCCAGCCAGGGGGAAGAGACATTTGCCGACCAAAAGTTTGTACCGAGAGCGCGGATT |-----|
94149
94159 AATTCAGAAAGCGTACCAACCGGGCATTAACCATGAGCAAAAAGGCTTTTGAGTAATAGA
94219 CTGATGACGATMAAAGGAGACCAATACAGCATCGCGCGCGTGGACGAGGTAATCTGA |-----|
94224 94233
94279 CCTGTGCA
...
94300 AGATTCAATCAGACCGGCAATCGGTAGCGACCTTCGATACCTCCCGGCTCTGGGACAAAT |-----|
94350
94360 ATCGTACTACCTGCGCTTTGTAATAGCGCACCAATACCAAGAAGTGAATGCAATATAA
94420 ATATTATTTGCCAAGCAAGCAATAAATGGGTGCTTTAGCGCAGGTGCTGCGATTGTGCAG |-----|
94447
```

```
...
90008 TGGCTGCCAGGCGGTGGTTCACATGCTTTTATTCACAGGGGCAACAGGGCGGACACA |-----|
90058 90061
90068 GTCTCAACCCCTCAGGCGTATCAGATPACGTTATTCCTTGTCCGGCGCTGCAGAGCGGGG |-----|
90094
90128 TGATCTTGAGATATCGTAAAGCTACGGCTGCTGCCAGGGCAATTGCCAGCCCA |-----|
90148
90188 GAAACGGCACCCAGAAAGTAAATAGTCACCGGGGCTGTACGGTATAGAGCGTATAGTTC |-----|
90219
90248 CCGACGAGCGGGGATTTGCTGTCTCATCCACTCGTACAAATCAGGGTCCAGTTTGACC |-----|
90312
90308 GGTAAAGGACCAATATATTAGATTGATGTAAGTCTGGCGCACGGATTTAATCGTTGCC |-----|
90359
90368 GCCACATCGTCGTACAGCTTCTGCTCCAGGAAGCAAGTGTGCCATGACATGTTGGTAG |-----|
90373
90376
90428 TCTTTAAGGCGCTTACGTTTCACCGGGAGTCGGTCAAGAGGACCGACCGCACAGAGG |-----|
90434
90488 ---|
CGATCGGCGAGTTCGGGTTGTACACTGAGGCAACATTTCCACGGGGCCA
90491
...
90623 ACGCTGGCAATGAATACGACCTGTAATTTATTCATGSGTGGGTTTAAAGAGTCATTT |-----|
90673
90683 CTAGGGGAATACCCCTGGCGCTAAGCATTTGCCACTTTGACGAAAC |-----|
...
90791 TCACCTCTTCCTGCCAGACATTACAGACACATCGGACCGCAATATGCGCTCTCTA |-----|
90841
90851 ATCCAGCAAAAGAAAAAGTGAACCTCAGGGTCCGAGAGCAAAATG |-----|
...
90978 GGAATTTGTCACCTCCAGCAGATTATTATTAAGTTTGACGGTCTTTAAAGAGTCATTT |-----|
91028
91038 GTGGGTAACCAAGCATGGCTTATCTGGGATAACCAATTGGCTTTGTTCTTCTATATA |-----|
91053
91098 TCCAGTACCA |-----|
91109 CAAAGATTGTTGGCCCTTCATTAAAGCGTAATATCATCGTGAATAAGGCACGCGTGA |-----|
91159
```

```
86901 TGATTCAATACCCCTCAGCGTTACCCGCGCTTCGCCGACAGGTGATACCCGTGGGTGCATT      |-----| |-----|
      86915      86925
86961 TTCACGGTGCCTCTGTTTACCACCTTCACCACATCAGAAAGACTTTCAGATCAATCGTTGGCC      |-----|
      87011
87021 GGAATTAATAAATAATCTCCCGGGTGACAGATTGTTCTGGCGCAATGCGCAGCGTGC      |-----|
      87068
87081 ATCAGCGCATAGAGGTTCACCTTTGGTCTTTTACGCCATTCTGGGTGAGCACCACATTTGGCC      |-----|
87141 CAGTCTGCATCTGGCGTGCAGACCGCCCGCGGTAAATCGATCCATGATGATGTCAGAGGA      |-----|
      87170
87201 ATATTGTAATCGGCTGCTGACCCGATTTGCTGACTTCACCGGTGACGTACGCTTTTTC      |-----|
87261 GAGCGGAACGCCGCGACGCTGACATCGACTTGGGGCTTTCATGACTGAATCCAGACGC      |-----|
      87267      87275
87321 GCTGTAACTCTGTTGCTGCTACCTGGGTAAAGTGTTCCTCCGGCCACTTTCAGCGCAGCAATA      |-----|
      87370
87381 TAGGGGTAGAAAATCGCACCATCGCGCTTAAACCCAGTTGCGGTATCTACTCGCGTACGG      |-----|
      87404      87431
87441 TATTGCCCTGCGGGGGGTGTTAGTTCGGGTGATCCAGACCGTAA      |-----|
...
88013 AGATACGTTACTGATTGCCGCCCATACCCCTACCGCCAGCGCGCTGTCAGGCGTTTGGC      |-----| |-----|
      88063      88070
88073 GAGCTCGATCGTTGCCCCCTTATCTTTTGGCAGATATTGCCACAGCTAACTTTCATCAG      |-----|
      88075      88085      88115
88133 TACGCCATTGAGCGTCGGCGGGGTTCCAGTAAAGCCGTCAAGAACCAAGTTGGCGTGAATA      |-----|
      88143
88193 ATCGGTGAAGCGCATGATTGTTCAGTCAAGTCAAGCTGCTTACGTAGTTAAAGTCCACACC      |-----|
      88241
88253 CAGGCCCCAGCTGSCATCTAGCGGCGATTAAGCAGCTCGGAACCGACCGGCATATCAT      |-----|
      88254      88276      88311
88313 CGTTCCAGATTAACCGCATACACTTGCCCATAGAAACATTTGGCTAAGTGGCAAGTA      |-----|
      88316
88373 ATTGCTTGCAAGTTGTTGAGATTAACGTCAATTGCGAAGTAATCACGATATCGTGGC      |-----|
      88397
88433 CACGCGGGGCGAGGTAGAGT      |-----|
...
88500 TTGGTGAATAATACCGCCATCAAGCAGCAGGTGTCGGTAAACACAGTAGCGGCGCTGGAC      |-----|
```

```
88560 ATCAGCCCCAGGTGACATATAGAAATCTTCGCCGGCCGCCAGGACTGTGAAAGTGT      |-----|
      88602
88620 GGGTTGAAGAAGTAGCTAAACGATCTTCACGAATACGGTAGCCCCGACCAAGCAGAA      |-----|
      88639
88680 AGATCTTCAGCTTCAACACGTTGTTGTGCAGTGGCTCTGATTGACAGCGGCTGTACT      |-----|
      88728
88740 GCCAGCTGTGTTGGCAGGCTGGCTACGTCGGTTTCGGTAGTCACCACCATATAGCTCG      |-----|
      88785
88800 CGCTTTGCGTACAGCTAATCTTCTCAAGCGCTTGGCGCAGGTTATTCACCAGAAATCCGG      |-----|
      88826
88860 TTGGCAGCATGACCCGCTTACAGAGAGTCACGGTATTTGTATTGCTGACAGCATATAC      |-----|
88920 AGCGTCTTATCGCGTAGCTGAATTTCTGGCGCTCAAAAGCCCGGTTATACTCAG      |-----|
      88920      88920
...
89025 GGCAGAGTTGATATGCGCGTATTGCGGTATGCGCAGAGGAGGCGCAGATCGTTGAAA      |-----|
      89075
89085 TTGTCGCTAAGTAAGTAAACCAACATCAAGCTGTTACCGGCTTCAATCAAGTTGAGG      |-----|
89145 TCTGCCAGCTGGCAGCGCGATTAACCTGCACCGAGCTTGAANTGGCTTGCTGAGCAGT      |-----|
      89177
89205 TTGCGAGGAATCATCTGTTGTTGTTGCGCGTCTGATTCGAGTTTCAGACGAGGGA      |-----|
89265 TTCACGCGCTTTGATATCAATGCCGCCAAAGATGGAAGCGGCGCAGAAAGATATCG      |-----|
      89271      89321
89325 CTTAAGCTGATATGCGCCCGCATCTGTGAGACTCTCGCGATGACAATTTATTCGCTCAC      |-----|
      89326      89336      89375
89385 CGGCAAAACGGGTGTTAATATTGCCCGCATTTACGGCGTACGCC      |-----|
...
89479 CCAGATACCTACCGTCAAAACAGACCCGTACAGCAATATCACGTTTACCAAAACGCACTT      |-----|
      89529
89539 GCGGTAGCAGTAACTCTTCCACAGGGAAGTTTAAATGCAATGATTTTGTCTTTGT      |-----|
89599 ATGACTGATCGCGCGGTGAAAATCTTCCACGTGGCTATATTACGTGTCGACATCCGTAT      |-----|
      89606
89659 AACGAATGTGTCTTCCAGCATGGAGAACCGCCACAGGAGGTGAGTAGAAACGTTATT      |-----|
      89688
89719 GATCGTTATCCCGGTAGTTGACGCTGAATTACCTTCGCGTGGCATGCGGCATTTGGCA      |-----|
      89720      89738
89779 TCTGCAACAATCTCG      |-----|
```

```
84516 GCTCCAGAGTTGAACGTGACAAAGCTTTTACCACTGCTGCGCTGCGACCGGTGATCATC |-----|
84537 |-----|
84555
84576 AGAATGTTATTCAGTCTCCATCATAGCAAAATG
...
84647 AGAATCCCGCGGGTTATCCACCCGAGAAAGGATATTTTAACTAGCATGCGCTGCTG |-----|
84697
84707 ATTGAAAAATTAATTTTCTTACGACAGAGGTAAGTTATTCAGCCACTGCGACATCGG
84767 GATGCTGGCATMAACGCTGATGCCGTCTCTTCAGTTGTTCCGGGGCTTCTACACACG |-----|
84779
84827 ACGCAACA
...
84850 CGGCACCCACCGAAATTAACAGCGCCAGAAATAAACACACCATTCACGCTTTT |-----|
84900
84910 TCGGTTCACTGGCTGGGGCTGAGTGAATGCCGGCTCGATTAATCCG
...
85055 ACGCAACACTTCCGTGTGGGTGGAAGGCATAGCCGATACCCGTTATTACAGGCTTGGC |-----|
85105
85115 TTCTTGCTCCAGCGCTCGGGTTTTCACAGACGCAAGATAAGTTGGTGATCTTCTT |-----|
85132
85175 ATACAGCTGGGAGATCTCTGCCCTCGCGAAGGTCAGCTCATTTGAGTTGATTATCAAGTT |-----|
85229
85235 CACAATCTGCTCAAGAGCGCTTGGCTTCCAGGTTAAGTCAACCGAATCGCGCTGCTG |-----|
85294
85295 GCGATTAACGCTGAGTTTCTTCCGCTGTGTCAGCTCGCTGGCACATTCTGGCAACTG |-----|
85299 85306
85354
85355 GCGCTGAAGAAATTCAAAGCTTGTGATCTCGCGCCCTGGCGAGCAATTTCTGTGG |-----|
85359
85415 CAATAGTTGTAGCGATGCTGTTCCAGATACGATACGATCACTGGGATCATCACAGT |-----|
85427
85475 CATGTATA
...
85495 CGCTTCTTACGCGTTCGTTACGTTAAAGTTTCCGCAACGCGTTAATCGCTTCA |-----|
85545
85555 GTTCGGTAGCTGGCTCAGGACAACGTGTTCTGCTGTGGCTTAATATCCGCGATAG
85615 TCAGCGCAACGCGCATCTTTTTCAGAGCGTGGCGGACATACCATTTGACGCTGAATCTT |-----|
85622 85639
```

```
85675 CACCTTCCAGCGTATAGTGG
...
85714 GTGTCACTGCTGATCTCTGACCATTCAGTTGTGAAATATGATCCAGCTGATCCGCACT |-----|
85764
85774 CACTGCTTCTCTTTGTTAATCTGGCCGACGCCGACCCACAAAT
...
85873 TACCGAATATCATCGCGATTTGACAGATTGATCTCCGTTGACACTCGGCGATGAGT |-----|
85923
85933 TAGGGATCAATATCACTCAGGCGCGCTGAGAAATGGCGTTGCCCTTTTGGTCAACCTGA |-----|
85964
85993 CCAGATATCTGCCGTGATAAATTGGTGTCTTAACAGCAGATAAGCGACGCGATCAAGC |-----|
86036 86042
86053 TGATTAACGGCGTACGCTGATATAATACTTAACGTTACGTTATCCATA
...
86157 TTCATATTTTATGTTGATCGTATTTTATAAATACTCAAAATACCGCTGAGGCGATTCGC |-----|
86207
86217 CCATTCCTGACTGCGCGCTCCACAACTTCCGTAGACTGTTCAAT
...
86299 TTCTGTTCAAGCCATTGCCCAACAGCATTTGTCTTCCCGGGACTCGGAGCGATTTGCG |-----|
86349
86359 GTAACGTAGCAATATGTCCGACTCCATTGCCAATATCAGATGT
...
86502 GTGTTTACAGACCATGAACACCTGCGGATTTGACTTTTACGCCCGGACAGCTTTACG |-----|
86512
86562 CAGCAAGCGCTCCGCAATGGGCGACGCAAAATGTTGCCGTTACAAACCAACAGGATTGA |-----|
86567
86622 G
...
86638 GCCATGTTGAATCAACCGTGAAGTCTGTGCAAGTCATCACCCCGAAATGCTCGGTA |-----|
86688
86698 CCAAGCAGGAATACACGCTTCCAAAGACGACGAGCGCGGTAGT
...
86781 GTTCCAGTATCATTTGCCATGGCTCCTTGGCATTCAGTGAATTAATTTGGGATTTTG |-----|
86831
86841 CCATTTCTGTTGTTTGGCTTGCACGAATTAACAAAATCCCGTCCGATCAGGCAATCC |-----|
86891
```

```
81754 CCTTATTTGCGGATTCGTCAGCCTTGGGATGGTATTTTGCAGAGTGGGCTCTGGTT
      81803
81814 ACAACTGAACACTGTTGGTGTGATTCATCTGCGTTCCACAGCTGTC
      ...
81980 ATGAACCTGGTGGCACTGCTACCTGGTGGCTGGATGAATATTTTGTGAGTCGCCGTT
      82030
82040 TTGTGATCTTCCCGCTGCTGGGGTTCCTCTGCCCCATTGCTACGCGTGAATGGCATTTAT
      82090
82100 CGTGGCCGCCCGGTTGGGATTTTGTGATGGCGTCAATGATGCAATTTGGCGATTTTC
      82129
82160 ACGGACGGCATCAGCGCTGTCCCTTTGTCTCATGCCGTCAGCGTGCATCTCCACG
      82222
82220 CTGACGTTGTGGGAGACGACTTCCAGTCAGCTGACGCTGAGTATTTATGTTGTAATCGTG
      82252
82280 CTGATATTTTGGCCCATTTGTGTCTCTACACTCTCTGTGAGCTACTACAAAATGTGGGG
      82345
82340 CGCATGAGACAGAAACTCTCCGCCGTAACGAAACGAGCTGTACTAAGAGACAGAAACA
      82400
      A
      ...
82469 CTGACCCGCCGATTGAAAAAGTTAAGAGACGTAAGTCTGATGGCGGCATTTACGCTGAT
      82519
82529 CAGGCAATCAATTAATGTGGAATGAAAAAGCAAAACATATGATGAAGCGCATCTTGAT
      82577
82589 CCCATTTTATCTCTTCTGATTCGTTAACCCCGCAATCTGCAT
      ...
82637 TCAGAGTGAGCCGGAGCCGGAGCTGAGCTGGAAGTGTGTTGATTTGACGTCGTCATGG
      82687
82697 TGTGCGTCCCCAACCAAGCCACGCAACTGATGAGAGATGTCACC
      ...
82756 CAAACTGGCCGGTAACACTGGGTTGGTGAACACCGCGGTGTGAGCTAATCGCTTATC
      82806
82816 TCGGACATTACCAAGCCAGCGTCTGTTAGCCGAGGATTTGCTGACGAAAAAGGCTGCC
      82828
      82834
82876 CGCAGCCTGCTAG
      ...
82984 CCCAGGAGATACGTCCAGTCCGATCCGTTATTTAATCTCTAATAAAACTGGCGTTTCCC
      83034

83044 AACTGATAACGCAACGTGACTGACGCGATCCTCAGCAGGGCAGAGGTCATTTCTTG
      83104
      ACTTTACGGGCATCGGCAAAACGCGCTTTCGGCACTGAGACGGGTAAATTTCCCGC
      83126
83164 AATCAAACTTGTGCTTA
      ...
83194 AGGAGAAAGCTGTTTATTAAAGCAGGCAATTACATCGAACTCAAGTTAGCGCGACGA
      83244
83254 ATGTCTAATTAACCGGTGCGGTAAGCCTGCAATCAATGCTGACGGA
      ...
83424 CCAGAGTTGCCCGCAGCCGCCACCCCGTTGTGGATTTGATATCATGATTAAGGTTGACT
      83474
83484 CCCCATCCACCGCAAAACAGCGGTATGTTGACATTACCACTTCACTGCTTTTATC
      83541
83544 GCGGACACGATACTAATCTGGCAATCTCGGGGGCAGCTGAGCTCAACTGAGCGCTT
      83546
      83597
83604 CCAGTCAAGCCGATTAACAGCCGCCAGGTGGTGAACGTGTTTGAACGCTGGCGTCGG
      83650
      83656
83664 CTACGGAATTAACAGCCAGTGTGATTCGTTGCTGTCTCCAGACTTTACAGAGAT
      83667
      ...
83909 CTCTAGGCTGATGACAAACGAAGAATGTTAATGCGTAGACCCGAAAAAGCGTTCAAG
      83959
83969 CCGCATCCGCTACCTTTAACCCTTCACTTTCTTCGAGTAA
      ...
84048 GCGTGTGGCGCGTTTAATCACACCATTGAGGATTCGCCCTTAATTAATTAACGCGAGCC
      84098
84108 TGTTCAGACGCTGATTGACAAAGCTCACTCTTTGGGGTGTTCAGGCCAAACGCGCA
      84115
84168 ACCACAGGCTGTGCCGACAGAGCGCCACAGACCGACGACTGCGTCAACCGCACATC
      84190
      84209
84228 GCGGCGTATCAACAATCACAGATGTAATGTGTGCGCCATTTCCAGTAATTGAGGC
      84260
84288 ATCCGATCGCGCATCAGCATTTAGATG
      ...
84456 TAACGAGGCGGTAATCGGCATCAATAATAACTTTTGTGAGTGGAGTGGGCGCATACC
      84506
```

```

79958 TATCAATACGCTTCGGGGAGAGTGTAAGTGTCACTGGCCATTTCAGCAGCGCTGACGACAG |-----|
79018 CGAAAGTAATACAGAGCGGATCTTCTGGCGATTATGGCGGTGGCCAGACGCTCGCGG |-----|
79078 CGAAAGTTGCTGAGAGACAACCTGGAGCGTGGCTGGCGCGCGCTGGCGGTGTGGCAGGC |-----|
79138 GGCACGCAAAATCTCTTCCGACAGATTGCGTTTACCGC |-----|
79326 GGGCGGGAATATTCAGATTGCTACCAATTGGCTATGGCAGAGCTATATCAACGCGACAGG |-----|
79386 GGTACGCCATGTCTGGCATTTCAGCTGTACGAGACCTTAAAGG |-----|
79433 CCGTTACTGGAAGTTATGAATCTGCCCAATACCGGAAGTGTGCTGGCAGCGCCAGAA |-----|
79493 GATTTCGCACTCTGGCGCAGCGGCTTAGTGAGTATGTCACTGGCTGGCGGAAGCTGCA |-----|
79553 CCGACGTAAAAAGACGGTAAGTATCGCTTTCAGGCTTATGAATATCGCAATCGGCAATA |-----|
79613 CCTCGTGTCTAGAGTTT |-----|
79909 CCATGAGATTTCAGTTTGTACTAAGTGTCACTGTCTATTCATTCACAACTATGTGGCGATTATTT |-----|
79969 TTGGCGCACCGCTGGCGATGGAAGCATTAATGACCTTCTCTCCGAAATCCACCTTTGTG |-----|
80029 GGCCTTTCT |-----|
80043 GGTGGCAGCGGCTGAATTAATACAGCACCTGCTGTGACGTGGCTGGCGGCGATTGCGT |-----|
80103 TCAATCTCTCTGCAATTGTGATATGAAATGCCAACGCGTGTGATGC |-----|
80244 TTTCGACACACCGTAATAGCGGCTACGTGACCGGGGCCATGTTTATTATGGCGATTCAGC |-----|
80304 GCGTGTATTTTACTGCGCGGACGGAGCGGATGTGCGATTACGCTGCTTGGCATGCGT |-----|
80364 TCCGCTTCGCTACTCTGGCGATTATGCGTACCTGCAACTGAGAGACAGATTCTGCGAT |-----|
80424 GAACTCCGCGCAAGTACAAACCGTAAAACTGGCGGATGGAAGGGAGCTGGCAACGGA |-----|
80456 |-----|

80484 CCTGCTCTGCACCGCTTCCATGTGGTTGCCGTGGCGGAACAGATCAAGAGCGTAAAGCC |-----|
80544 TTTCGCAATCAAAATTCCTCCGTGCTGTAGGATCTCGCCACTCACTCAATT |-----|
80689 CGCAAGGCATATCGTAGCCGCATGTGTTCAGAGCATTCGGGAGCTGGAAGCGCACTGG |-----|
80749 GCTACGGCATCTTGCCTCTCCGCTATGGCCCGAATATGAATCATGACAGCCGACAGT |-----|
80809 ACCAGCGCGCATGTGGCGTGGCGGATACCTCAGGTTGGCGGCTATTCTGAGTTTCGCA |-----|
80869 TCA |-----|
80874 GTGGGCTGTGTTCCCTGCTGCTACTGTGATGTGATTCGCTTGTCCAGACGCTGCGT |-----|
80934 GGCAAAATCGACCAACATCGCTGGGTGCTGAAATATGGGCTTGGAGTTTGGCTTGGCG |-----|
80994 TGGATTGCGATTGAAGCCGGGTGTTATAGCCGAGTTTGGTGTACGCCGTGGGAGATA |-----|
81054 CAGGACATCTTACGACATPACTCCGCGCACTCCGCTTTAACCAAGAC |-----|
81103 AACTGGCTTTCCTCACTGATCATGATGTAGGGCTTTACACCCGTCTTAAATCGCGAAG |-----|
81163 TCTACCTGATGAGAAATATGCCCCGTGTGGGCCGAGCGCGATGCA |-----|
81334 TGACATGGGGAATTGGCTGTTTGTCTGCCCTGTGTGGCGGTGATGATGAGACGCCGAT |-----|
81394 AGTGATTAATAGCGTTGGTGGCGCACTGGGAAGCAATCAGTCTGTGATCCTCGCAGG |-----|
81454 TGGGCAATTATTTGCCCGCTGGCCAGAGTATATCAGCGCGGTTTCCGGCTTTATGT |-----|
81514 GCGCATGATCCTGGTGTCTGTCTGCTGCTTCTTCCGCCCGCTGGCCTTTGATTATCGCGG |-----|
81574 AAAAATGCGCGATGACAGCTGGCGTAAATATGAGGATGCGGCTGTGATCATGCGAGTCT |-----|
81634 GGTGCCCGCGGTAGTCTTTCGATATGCGCTTGGCAACTTGTGTCTGCGCGTGGCGCTTTCG |-----|
81694 CTTCACTCCGAATTAACGCTGTGAGTATCTGGCAGCTTCTGGCAACTGTGACGCCATT |-----|
81745 |-----|

```


...
76676 TGGGAACCCACACCTGGCCGACAGAGTGCCGTGCTGTTTACCGAAACGCCGCC
76736 |-----|
GGGGCGTTAGGCGACACTGGGGCCGCATTCGCATGGCAAGATTGATCTCTACCAGTGCCTG
76738 |-----|
GTGCCGACCACTGGAAAGCGCCGCGCATCCCAAAGGAGAGATTTGGCGCTTATGAA
76856 |-----|
GCGGCGTGTGATGACCAACCAAAATGGCGGATCCCGCAGCAACCGCTGGAGATCCTCGGTACT
76858 |-----|
76916 CTGCACACCTTTGATCCGCTGCCCTGCTTCAACACACGTGCTGGAGACGGCTAGC
76916 |-----|
76961 76964
76976 GAGCTGATCTCCGTGACAGGTGCGTTAACAGGAGAGAAATCATCATCACAAGAAAGC
77036 |-----|
77038 77065
77039
77096 GTGTTATGATGCGCGGTGTGATGGTGACCGGTTACTTTATCGGCMAACCCGCTACTTCC
77156 |-----|
GTCAGCGGAGGCGGAGGTATCTGTTCATATATGGGCTACATCAGGTTAATTCACCTCAGC
77216 |-----|
77218
...
77320 CTGGTGACAGGGGTGTGTATGAAATCCGCTGSTATCTGTTCTGGCAAAAGCCGAG
77380 |-----|
TGCCGATATAGCCCAATACCCCATTCGCCGAGCGGATGTTGGCTATTTCTCTGATGTC
77440 |-----|
GGTCTTATGATCATCACCGGCTTTGGCG
...
77573 CGGCTGGGATGTGCTGATTTGGCGCTTGTGATCGTCATGTCATATATGCGCTCGCT
77633 |-----|
GAAGACATCATGTCCGACGACAGGTGATCTCCACCATGTGTCAACGGCTACCGTAGCCAC
77648 |-----|
77693 AAATTGGCAAAATATAGTACAGAGAGCGTTTATGAGCGAGCAACGCGTGTGTAATGG
77718 |-----|
77753 GCGTGCGCAACCTGCTGTGGG

...
77843 AGATTTCGANTGCGGTACTACAGGACTGAACTTCTGGGATATGTCAAAACGCCAGCC
77903 |-----|
ATCTGTGATTTCTGATGCCATTGACTACGGGCTGGAAACAGAAAGCGTGGCAACTATG
77963 |-----|
CCGAGAAAGTATTCGGCTTATCTCAGCGCGAAGAAATGAGCCTGCACCGAAGACGT
78023 |-----|
TCTCCGAAGTGTGCGCGCTGGCAGATATCCGGACATCTGCCACATATTTGCCCTCG
78083 |-----|
TGGTGTGCAACCCGCATATGCTCGACGACTACGGGGCGACGCTGAGCGAATCTGGCAGGG
78143 |-----|
AGCAACTGCCCGCAGCGGAACAGCGCGCGCTGGCGCAGCTGCTGATGGGGAATTGTC
78203 |-----|
CGCAACCGGCTAATGATCGCGCTGTCTCAATTATGACTGTCTGTGATGGAATAATTACG
78263 |-----|
AAGCGTTTCGCTTGGCCCGAGTACAGATGACACTGGAGAGAGGAGATGAGCAACGACAC
78323 |-----|
GCCATTGATGCGTTGTGGCAACGAATCTGGCGCGGCTGGAGCGCGGTTAGTGAATT
78383 |-----|
CCGTCTGAGGACTGCGTTACGCAAGCCGACAGCGGCTGTGTTATTAAACAGTAGCC
78443 |-----|
GAAACGACGCCAGAGTACAGGATTAATCCGTAATGATTGGCGAATTACTGCGGAGTT
78503 |-----|
TCCGACTACACATGCGAGGTGGCGATTGCTGACCTTGAAACAGAGCAAGCCATCGAGA
78563 |-----|
TGGCTTGGCGCTCTTGGCTTTCCCGCACATTAGTGTTTACGGCGGAAACTATCGCGG
78623 |-----|
CGT
...
78699 GCTTCATGAGGAAACTTTTTTCATCTGCTGGGGCCAGAAACCAACCGCATGACA
78759 |-----|
GTTTCACCATGAATCCATGCCGATCACCTGTCTGAGTGAATGATGA
...
78838 CAGCCCGCAGAGTATTCGCTGTTAACGAGTTACAAACATCACTAAGTCAACGCCAGCC
78898 |-----|
GCCGTTGGCGAGGTGCTGCGACTGATCTGTAAATCTCAACGCCGACGATGCTACTT
78938 |-----|
78944 |-----|
|-----|

```
74063 GCGGATGTGCAAGAGAAAGACGTGCGATATGAATTAACGAGAAACATTTTACCAGGC
      |-----| |-----|
74123 CATGCGCGCTCAGGCGGTACCCGCGGACGCTTTCATAATATTGATGCTGGCTGCAC
      74128 74136
74183 |-----| |-----|
      GTCGCTGGGATTAAGCGCGGAATGGCACCACCAAGATTGCTGGCGTGGAGAACAAC
      74183 74225
74243 GCGCATCCGGTGTATGATCCAGGCTCGAATGCACCTGCTGACGAAATCTTTAT
      |-----| |-----|
74303 CCGCTCCGTCACCCCTGGCAAGAGCGTCATTCTTCCGTATTTCCCTGCATTACGA
      74328 74361
      ---|
74363 CGATACCTTTGANTGGCTGCCGCGGAACCGAGCGGAAGAAGTCTTGAAGACAT
      74366
      ...
74506 GACCGTTTATGAGAAACTCAAAAGTGCCTGCCGTGCCAGCCGACCTTAATCCGACACAGGCAACGT
      |-----| |-----|
      74556 74621
74566 GAACCTGTGCTGCTGGGGCTGCGTGCAGGCCGCGGACCTTAATCCGACACAGGCAACGT
      |-----| |-----|
      74621 74626
74626 |-----|
      CTATCGACAAAGTCATCACCGAACAACCCATTATCAAAAGTACCTGGCTGCCCGCATCC
      74626
74686 CCGATGTGATGAGCGGCATCATTTACTTACATGATGACCTTGTATGCTGCCAGATGTCG
      |-----| |-----|
      74697 74728
74746 ACAGATGGGTGTCACATGATGTTTATGTCAGCGTATCCAGATTAATGCTATGCTC
      |-----| |-----|
      74755 74800
74806 GCGTCACCTTCAGCGCGGAGAGATTGTCAGAGTTGGATGATGAGCGCTGCCGCAAG
      |-----| |-----|
      74817 74850
74866 GTTATTGCCGTGACAAATGGCTGCAAGGGCTTACACATATTAAGCGCTGTTCCGCA
      |-----| |-----|
      74910
74926 CACGCTGGAATGATGGCGTTCTTCCCAATCCAGTCTGTACGGCTGCCCTGGCTGTG
      |-----| |-----|
      74940
74986 CGGAAATG
      ...
75073 TCGTTTAACCGCGCTTGCTGTGTTGGCAGCGGCTGTGTTGTCAGCAGCTGCCAGCG
      |-----| |-----|
      75123 75129
      -|
75133 CCGTTGACCAGCGCAGACCTCATTAATCAGCAACCTACAGAAACCAATCAGCCAGGCA
      |-----| |-----|
      75134 75148
75193 ATGAGGATAA
      ...
75205 AGGCATGAGCAGCTCAGTACGAACACTCAGGATACACCATCATATATGCCGAGCCGCT
      |-----| |-----|
      75255 75255
75265 GGTGTGACACCAGATTACGGCATTTGAAGCCACATGCGCTGCCAAGTGAATATTAAAG
      |-----| |-----|
      75320 75320
75325 TCAGATGTATATACCAATGCCGTCTCTCCGCGACCATGTTTCCGCGGCTGGAATCAT
      |-----| |-----|
      75325 75325
75385 CTTGCAAGGGCGGACCCCGCGCATGCGTGGCGTTGTTGAAGCATCTGCGGCGCTG
      |-----| |-----|
      75415 75437
75445 TACTGGCGTACAGCCTCTGCTTCGTTTACGCCATCGAAGATGCCATCGGTATTAAAGT
      |-----| |-----|
      75505 75544
75505 GCCGGACAAGCCCAATATCATCCGACACATTATGCTGGCAAGCTCTGTGCCAGATCA
      |-----| |-----|
      75512 75544
75565 TCTGTGCACTTCTATCAGCTTCCGCGGATGACT
      ...
75922 TTGCGCCATTACATGACGAAAGCGCGGTGCGGCGGTCAATATGGAACGCTGAA
      |-----| |-----|
      75972 75972
75982 CTTGTGCACTCGATTATACCCGCGACGCACTTATTAAACAGTGATGATCCCGCA
      |-----| |-----|
      76040 76040
76042 CCGCTTAGCCATCGGTGATTTCAACAAACCGTGAGCGGAATGCGCACGGGTCTTCTGA
      |-----| |-----|
      76045 76045
76102 TAAATGTGTCCTACGTACGCGCANTCCCGATATTGTCACAGCACTTGTGTGAGAAAG
      |-----| |-----|
      76142 76142
76162 TCTGTGATGCTGTGGCGCGGTGATTAAAGCGACTTCAACAATGTGCTGCCATGGA
      |-----| |-----|
      76193 76193
76222 TTTGTTGATCCGACAGGTGACAGAGTTTGTGATCAGCGCTGTATGTTATCCCAA
      |-----| |-----|
      76269 76280
76282 CGATCAGGTGCGGCGTATCCGTTGATGGCATACCGACCCGTGTACAAACCCGCGCA
      |-----| |-----|
      76285 76293 76337
76342 TGTCAAGGAGGATACCAACATTCAGCAGCTGAATGAACAGGAAGCTATCTGTGAT
      |-----| |-----|
      76342 76352 76386
76402 CAAAGCGCGCGCTGGCGCGGTAAACGAGTGAAGTGGGCGCGCTGACAGCAAGTTAAT
      |-----| |-----|
      76405 76460
76462 CCGTTATCAAAAGGAGATGCTGCAACCGTTGATGCTGTGATGATGATGCTGCGGCT
      |-----| |-----|
      76465 76475 76516
76522 GAACCTGCCGCTTCCGGTATCCAGTCAAGCTTGGGCGCATTTTGTGCCGCCGACGA
      |-----| |-----|
      76549 76549
76582 AGCACAAGTGGCGCGCAGTAAAT
```

```
70534 TTGAGTACGGTCAGTACAGCGTTATTCTGTGTGTGGAAGGTTCCCGCCGTCACATGCCG
70535 70551
70594 GGACCATCACCGGTATGAAGATCTCAACCGGGAGCGCTGATGATGTTTTCTCGGTGCA
70628
70654 TGTCCGAGATACGTCCTCCGCGCGGAGGACGCTGCTGCTTTGTAAGTATGTTGGAAGANG
70687
70714 CGGCGGTCACGCTGAGAGGGGCAAGAAAGATCCGGAGAGCGGAGAGCTCCGGAGGA
70760
70774 ATCCGGCATATCACCGACAGTCAGGCAAGAGAGCGCGGCAA
...
70848 TCAGCCCGGCAAGCGGCAAGAAAGTCAGCCGCTGCAGAAAGCAGTCAGAGGAGGCGCTCTCG
70898
70908 TCTCGGCTTCTGCGGCGCTCAAAAAGCCAGTGAATCATCAAAA
...
70971 TTGTCAAGAAAGACGCGAAGAAAGTCAGCCGTAATGCAGCCAGGAGTCAACAGACGGA
71021
71031 ACAGAAAAAGCCCGGAGTCAAGCAGAAAGCCACAGTCAGCGGAAC
...
71149 AACCGGTCGCCGCGGCTCCTCAGGGGCGCAAGGAGATAAAGAGAGCGGTGGCGACACCG
71199
71209 GCCCGGCAAGGGCAACCGGCGAAGCGGAGCGGAGGTGATCTGG
...
71392 AAACAGGTGCGGCTGGCCCGGTGGGGGCAACCGGACTCAGGGACCGAAGCGGACCGG
71442
71452 GGGAGACACAAATCCGTTTTCTGTGGGGCGGCGAGCATTTATGA
...
71667 GATACAGGCAGAACAGAGAAATGACATGACATATTAAAAAAATTTATGACGCGTCTG
71717
71727 CGGTTGCGGAAAGCATGATGACTGTGAACAGGGGCAATCTTACAGTACAACTGCACT
71763
71787 GGGCGGCGAGACATCTGAGTCAATGAGA
...
72041 CCGGTAACTGTACAGTACAGCTAATATAGTAATTATATATATATACAGAGCGCTTGA
72091
72101 TGTATTATGCCATGCTCTGCAAGATGCTGCATCAGACGCTGAGCCACATCAGGCAAGG
72136
72161 TCTTGATTGTTCATTTTGTACTGGGCGGCTTGGCGGAGCCAGTTAGGTGCCGAGAGAT
72186
72221 ATGTCAGCCATATTCGGGCTGTTGAATGCGATGCTCTGCAAGATGCTGCATCAGAGC
72281
72281 CTGAGCCACATCAGGCAAGAGCTTGTGATTTGTTGATTTTGTACTGGCGGCTTGGCGGAG
72282
72341 CCAGTTAGGTGCCGAGAGAAATATGTTACAGCCATATTTCCGGCTGGTGTGATGCCATGCTC
72401
72401 TGCAGATGCTGCATCAGACGCTGAGCCACATCAGGCAAGAGCTTGTGATTTGTTCAATTTG
72418
72461 TACTGGCGGCTTGGCGGAGCCAGTGAAGTCCGGAGGAATATGTTACGCCATATTCGG
72468
72521 GCGTGTGATGCAATGCTGCTGTGCAAGATGCTGCATCAGAGCGCTGAGCCACATCAGGAG
72559
72581 AGCTCTGATTGTTCAATTTTGTACTGGCGGCTG
...
72650 AGCCATCTCCGGGCTGTTGAATGCCATGCTCTGCAAGATGCTGCACAGACGCTGAGC
72700
72710 TACATCAGGTAGAGCGCGCATGTAAAGCGTTTCTTAATCTTGAATGAGTAAAGATGT
72734
72770 GGGACGTTGAAGTCTGATTTCTGTATTAAGAAATTTTAACTTAAAAATAGTTGTGCA
72772
...
73101 GATATTATTACGTTTTTGAAGTGTGTTCAATCAGCTATAGATGAGGCTTGGCGGACAGAA
73151
73161 TGGAGATATTTAACGAGGAATAGAAGCTTGTATCAGAGCGGCTT
...
73571 CTTGATTATTGCGGGAAGCGAGAGATTCGAATCTGGAACCTTTGGGGTGGCGGCTT
73621
73631 TTCAAGACCGGTGCTTCAACCGCTGGCCACACTTCCGGAATGAC
...
73823 ACTGCTAAACATCCCAACCGCCCGCAAAAGAAATAGTCCATTTTATTCGCTAAAA
73873
73883 GATAAATCCACAGATTTGTAATGTTTGTGCAAAAGTTTCACTACGCTTATTAACAAAT
73943
73943 ACTTTGCGCAGCTGGCGCAGTGCAGAGATGAGCTTTGCTTTTCAGCATCTCAGCTG
73950
74003 AAAGGATGTTTGGCTTGTCTACAGGAGAGCTGCTTGGCCAGCATTAAGCGCGGTGCTT
74004
74028
74031
```

```
67821 GGATGGCGTGGAGACATACGGTGTGTGGGGGCTCTCCCTGCCGTCACCTGCTCGCGCT 67821
67881 GTTCGCGTGTCTCTCCATCCGGAAACACGAGCGCACCTTTGCCATC 67874
...
67961 CCATCGTGATTAACGGTCCCGCTTTTGAAGCCGAGTCAGGACCCGTGAACAGCGTCATCC 68011
68021 CTCGGCAGTGCAGACACTGACGGTGGAGGTGAGGGCAGCTTGACGG
...
68148 CGGAGAAAGCAGCCGTCGTGTGACACCCGCATCACTCGGATACAGAGCAATCGTTCAG 68198
68208 TGGTCTGCCGCCGGGGGAATACACCTGACGCTCAGGGCGGATTAACAGTTATGGCCAGCA 68244
68268 GGGGGAACCGGCCACACACCTTCAGATTA
...
68486 TATCCGGCCCGCACATTAAAGCCCGGAAGATTTCTGGTTTACGTGGCGAGCGCAACC 68536
68546 TGTGGGGAATCTGCTTTTGTGGAAGCCAGTGGCCGCGCAGCAA
...
68707 AAGACTCTCATCACCGAACCCGCAATGAATACACACAGACGCTCAGTAAACGCTGAG 68757
68767 GACCAGAGCGGCATCATACAGCAGATACAGCGCGTGCAGAGAGACACAATGATGACCTT 68773
68827 GC
...
68975 ACGGCAACACCACGCCGTTATTGTGGCGCAGGGGAATCACTGTTTCAATGACGATGTGT 69025
69035 TCCGAAACGACCTGTTTGGCGTGAATACACGTCATCCGCCAATCCCCGAGCTTTTCCC 69041
69095 TGACGCGGAGAGGCGAGCTGACCCGAAAGAAATGCTGATATACGCGTAACGTGAATGCA 69096
69155 ATTCCGGAGCGCTCAACACGTCACGATTAAAGAACTGTGCGGTTCTGGAAACTGT 69172
69215 CCGGAAACCAAGATTGAAGGGGATCTCGTTAAACAGTGGGAAGCTTTCCCGCGGATTT 69232
69275 CCGGTGCACCGGAGCGGTGGCCATCAGGACATTAACGTCAGGGTTTATGACGATCAGC
69325
69335 CGTTTACCGCGCAGATTGTTATTCGGCGGTGGCATTCAGCGCGCTTAACATGAGAGAG 69376
69395 AGCACTAGATATTACTCTCTCATGCGCTGATAGT
...
69443 GTGCTGAATTTATTAACCGTACCGCGTGATTAATACGTGATTTACGTGCGCTTATTTG 69493
69503 ATATGCCCTGCCGTCACGCTCACATGACGCTGAGATTTCGGTGTGATGAAGAAAGCCA 69528
69563 ATTAAGTGTATCCACACAGATATACGCAATTTGTGTTGTGATGAAGAAAGCCA 69589
69623 CCGCAGGCATCAGTATCAGCTGAATTTTAAACCATTAACGGCGCCAGAAATGCGCC 69665
69683 TTTTTTATTCAGAAACGAGAGGTATTAATGGTAACCTTTATGCCCATTTTGTCC 69677
69743 GCAGCCATTTGCTGCGGCTATCCGGTCCGCTGACGCTGAGGCGCTGAAACATCAGTCAAG 69781
69803 CTGAGCGCGGGGTATCTTATGCTCTCAGCAGACGTTCTGCGCAGCGATGAACGAGCGG 69804
69863 ATTAACGTGAATACCGTTATGAGTTACGAGACACACTGGGAGTGTGAGCGTCTTCAGC 69910
69923 TATCAGAGAGACAAAGATCGCCAGCTGACCCGTTAACAGGATACCCGCTGCGATGAAGAT 69939
69983 TCCGTTGTAACCGCTGTTCAAGCTTAATGCG
...
70114 AACTGACAAACAAGGGGAACGCAATGATGTGCTACCCGGAAGTATGACGCTCCGACAG 70164
70174 CAACACGCTCTCTGCGGTGGGAGCTGGCGTGCACATTTAAACCGATGCAATCCGTGGCCAT 70262
70234 TGATATTGCTTATGAGGCTCCGCGAGTGGCGCAGCTGACGAGGCTTCAATCGTGG 70262
70294 TGTGCGTAAATTAATTCGATTAGC
...
70414 CAGGAAACCGGTAGAGAACTGCACCATTCACATGAAGCAGACGTAACGCGCCACGG 70464
70474 TGTGTGTAACACGGTGGCTCTGAATAATCCGGATGAAGCCGGTCTGTTAAGCATGAGAGC 70516
```

```
63966 CCATTAAATCACTTTCAGATAATCCGGTTTATGACACAGACTGATCAGGATTTG 64016 |-----|
64026 GCGCATGCCCGGGGTGTGCGCTCGGAGATGTCGGCTTGCTGTGAGACGGCGGAG 64076 |-----|
64086 GGGAGCGGTATATCCCTTGTGTGAATATCTCTGCAAGCGCGGAGG
...
64422 GAGGATGACTGTGTGCGCAATGTCAGAACTGTACTGTGACATATGAGGCGACTGGT 64472 |-----|
64482 TTTTACAGGATTTCCCTGCTTCGCGACAGCCTGGGATATCCCTGCTGTGCTTTGGC 64515 |-----|
64542 GCATCGGTGGCCAATCATGCCGCATATA 64545 |-----|
64595 CCATCTGCTGACAACTGAGTAACGGAGAGGTATTCCGAAAAATGGCAACGAC 64645 |-----|
64655 GCATTCAAGCTTGGCTGACCGCCACTGGCAGCTATCTGCTTCAAGGGGATTTACAA 64710 |-----|
64715 TTTGGCCGCGCGCTCAGCCTGTATGTGAACAGCGGAGGAGCCATCCGT 64715 |-----|
...
64955 CTGGGGGGGCGCAGCCATGCTGGGCTTTTCTTCACTGCCGGGCATCATGCGCTTATGG 65005 |-----|
65015 GGTTCAGCCCTGGCAGCCGCTGTTTCTGCAACACAGATGCTGT 65005 |-----|
...
65313 AGGCATAAAAGCGAAAAATCCCGAGTGTCTACGGACAGGAACTGCGGGAGCGTTACGA 65363 |-----|
65373 AGATTGAGTGTAAAGAAATTATTATTGTACAGCAAAAACATTAACTCAGAGAGGAGG 65363 |-----|
65433 ATGTGCGCTTCTTTTCAAGGAGAAAGATTTATGCTCCTGAGGAATAAAGSTAAGGGCC 65464 |-----|
65493 CGCCCTTACCTGACTGATTATGAAT 65464 |-----|
...
65556 GGGCTCCGATGTCATGATGTTATCACCGCCAGCATGAAGCATGAGAGACGCCCTTT 65606 |-----|
65616 AATCTCTTTTAATAGTTCACTCTCGGGGCGAGACCGGCTAGTTCCTTTTCCGTCATG 65606 |-----|
65676 CGTCACGAACAGCGCAGGAGGTGCGCCGGGTGTCATCCGAGACGAGGGCCAAAGATG 65697 |-----|
65736 TTTGCCGCTGTTTTTCG 65697 |-----|
...
66045 CTCATTAAATCCGTTAACAGACTGAGCTTGTCTGTTTACAGGGTAACAAACAGCTTCTGA 66095 |-----|
66105 TGATATCGCGCAATAGCTGTGCATATCTCATCTGCGATTATA 66095 |-----|
...
66503 GCCTCGGGTACCTTCGGGGTGCAGTCACTGTTGGAGACCACTCAAAAGGCGACGTA 66553 |-----|
66563 ATCCCTTCTGTCCGACTGCTGATTCAGTTGACCGTTAACGGTAA 66553 |-----|
...
66613 GTGACGGAAGAAGATGTCACCATTAACGGCAAGACCACTGCGAGTTTCTGGCTGCTG 66663 |-----|
66673 ATTCTGATTAATCTGCTCCCGTCTTTTAAATCCGATGCTCCGGAGACGGCGGAC 66663 |-----|
66733 AGCACTCGGACAGCTGCAATTAAGCAATTAAGACCTCTGTCTCATTAACCGAAATCATGAT 66759 66768 |-----|
66793 GTGAACAGTCTTACCCGAAACAGCGGATTTGGGGCTGCAAGTGGATGCGGACAGTTT 66816 |-----|
66853 GGTGGCCAGCAGATGACGG 66816 |-----|
...
66987 GGCCTGCTGCTGTGGGACATGCTGACTCACCAGCGCTACGGCATGGAAACCGCTGGG 67037 |-----|
67047 GCGCGGAGTGTGACAAGTGGCGCTGTATGCCGATTCGCGAGTACTGCGACGATGCT 67068 |-----|
67107 CCTGATGTTTGGGG 67068 |-----|
...
67130 AGCGCGGATGACCTTTAATGCTTACCTGTACACACAGCGTAAGCGGTGGAGCTTCA 67180 |-----|
67190 GTGATTTCTGCTCGCGATGCGCTGTATGCCGATATGGAACGGCAGACGCTGACCTTCG 67203 |-----|
67250 TTCAGACAGCCGCTCGAGATGTGTGCGCTTACACCAACAGTATGTGTGTGATG 67236 67242 |-----|
67310 ATAAGGGCGTGGGTTTCGCTACAGCTTCAGCGCCCTGAAGAGACCCACAGCGGCTGG 67339 |-----|
67370 AGGTGAATTACACCGACCGCAGAA 67339 |-----|
...
67571 GGTTCACAGGGGCTGCTCACAACCCGGTACATTTATGAATCTGTGATACGACTATG 67621 |-----|
67631 CCGGACCATGACCGCGGAGCTGTCTCTTCATCATGATGCCGCCAG 67621 |-----|
...
```

```
61201
61207
61233 GACCCGCGCTGTAAGAAAAACATGGGACACTGGAGACCTGGCGGATTAACCCGGGAA
        |-----|
        61238
61293 GGCATTCAAAATCGATGTGGATGCCATTCTGGATATCGGTGCTGCTGAGTCCTCAGCGGA
        |-----|
        61331
61353 TATGCTGCCAGTGCACAGAAAGCATTTGATGAGCGGATTAATAATGGCAGTGTACCA
        |-----|
        61424
61413 GAGCCGGACGACGGCGGGGAAAAACCGCCTCTTCCGGCCAACTTCAGGGCGCATG
        |-----|
        61473
61473 GAATGAC
        ...
61569 ACGGGAGCGCATCACAGCTGAAGTATACCGGAGAGCGCAGAGGCGTATGAGCGTCTGCT
        |-----|
        61619
61629 |-----|
        61629 GAGCCCGCTGGAGAAATATACCGCCCGCTCAGGAAGAACTGAATAAGCCCTGAAG
        ...
61699 CGGGCGGATTACAAACACGCTGATGGCGGCGGAAAAAGATTATGATCGACGCTGAAA
        |-----|
        61749
61759 AAGCCGAAGTCGTCAGGAGTCAAAAGTGTCAAGCCGGTGAGCGTCAAGGAAGACAGCGCAT
        |-----|
        61767
61819 GCTGCCCTGCTGGCGCTTGAACCGAGCTCAGAGCGCTGGAAAAACACAGCGGTGCAT
        |-----|
        61830
        61851
61879 GAGAAATCAGCCAGCAGCGTCGGGATTATGAAAGCGGAAATCAGTATGGGCTCTG
        |-----|
        61895
61939 AAAGAGGCTGCC
        ...
61954 AAACGGCAGTTATCTAGCAGAAAAATTCCTGCTGGCGCATTAAGACGAGACGCTGAG
        |-----|
        62004
62014 TACAACGGCAGCGTGGCTGAGCTGGGAGACAAGTTGACACCCAGAAACGCTTGAATGAG
        |-----|
        62018
        62060
62074 CTGGCACAGCAGCGGTGCGTTTGAAGAGCAGCAGAGCGCG
        ...
62136 AAAAGCCCGGGTCTCACTGACCGTCAGGCGCAGCGGAGTCTGAAGCGCAGCGCTTTCG
        |-----|
        62186
62196 GGAAGTGTACGGTATATCCGGCTGCGCTGGGAAGGCCACATCT
        ...
62353 GTTAAAGTCTGCACAGCAGACCTTTCAGCGTATTGACAGAAATATGGCGCATGCTG
        |-----|
```

```
62403
62413 ACCGTGCAGAGCAGACTGGCGGGGATTCACCCGTTCCGTGCTCCATGCTCAGAG
        |-----|
        62487
62473 ATTTTCTGAAGCAGCGCGATGGTGGGATAGTCCGGAGTATCGCAGCGCATTTGGCGGT
        |-----|
        62518
62533 GCTTTCGGTGGTGGTGGCTGCTGCTCCACGGGGAGCGGCAT
        ...
62651 TTGTTCAATCGCGGGAGTTTGTCTTCACGAAGAACAACACCGGATTTGGCTGGCA
        |-----|
        62701
62711 ACCTGTACCGCCTGATCGCGGCTATGCGGAAGTGGTTATGTGGG
        ...
62783 GCGCGGGGGAAGTATTATTTAATCAGAAACATCAAGTGTGATTCAGAGACGCGCA
        |-----|
        62833
62843 CCAACGACAGAGCGCGGCCGACGCTGATGAAGCGGTGTATGACAT
        ...
62907 GGATGAGCTCCGCTGAGTTGCGTGTATGCGGATGTATTATCGGGAGCGGCGATGAAA
        |-----|
        62957
62967 ACCTTTCCTGGAAGTGAAGCCGATATGAGAGTGAACCTGCACG
        ...
63048 TCACAGCGTATGCGCGCAGGCGTGAATGTCGACCTGAAAAACATPACCGTGTACGCTTCC
        |-----|
        63098
63108 GTGACCCGGAGAGAGGCCCGGCATCTGGAAGCGTTCTGSCAGACAGCAGGTGCTGAAG
        |-----|
        63137
        63167
63168 GCGTTTCTGTGAGACCGCCTTATGCTGCGCGCAGATAAAGTGAACCTGTGCCG
        |-----|
        63172
        ...
63470 ATTGACGGCAGCGGTTTTGAGATGACGGGAAGGGCAGCAGTGCACACCGCTGACG
        |-----|
        63520
63530 GTGTCAATCTGTTTGGTGTGTCACCGGATGCGGAGAGACTGC
        ...
63633 CGGTAAATTTTGTGCGGGCAATCCGGAAGCGGACCGGACAGAGACTGAGCGCGCT
        |-----|
        63683
63693 GGGTGTGAGCAGATGTCAAGAGCTGACGGCCATGACAGCCTGTTGTGTGCGCCACAC
        |-----|
        63769
63753 CGACGAGACGAGCAGCGCTTTTCCCGTGCATTAATGCTGGCGAATACCTGTATGT
        |-----|
        63769
63813 GGACCTACCGCT
        ...
        |-----|
```

```
57923 CTGTGCTGGCTGCGGATCAGGTAAGGAGAAACGGTCTCTGCATC
...
57988 GGCATCATGGCGAACCATTTCACCGGCTGGGTTACAGGGTGGGGTGCACACGACGTCAGCA
|-----|
58048 GACGGTGAATGGGACGTGAAAAATGCTCATGGCCATGCCATGAAGTC
58038
...
58189 GCTCATACAGCCACAGGACAGCGGGATTTCTCGGCAACCGGCACTGACGCCCTTG
|-----|
58249 ATGCTGATGAAGCCAGCGGAAACTGGTGTCTGGAGCGACAGA
58239
...
58411 GATGAACACAAAAAGCAAAATGCTTTGCGGCGAGTGCCTGATGTCACGCGCGCTGCCG
|-----|
58471 TTAACAGTTATCAGGCCACCGGGTGGCTGACTGATTTCTGAATG
58461
...
58621 ACGTGAAATTTCCATACCGAAGAGGTGATGCTGACAAAATTTACCGAAAACGCGGTG
|-----|
58681 GCGGCTATGTCCTCCCGTTGTTGAAGAAAAAGTGTGCTGCTATC
58671
...
58789 AGCCGGTAAAAAGAGCCGGTACCACTGTTGGTTTATAACGGTCAAGGTGACGGCTATG
|-----|
58849 CAACCCGTTGTGACAGCATGACTGGCAGCGACTGGCTAAGGTGAAGATCTGACGCCGG
|-----|
58863 58876 58901
58909 GCGAGATGACGGCAGAAATCTACGATGATTAACCTGATGATGATGAG
...
59023 GGAACCCGGAGAGAGAGGCCAGAAAAAGGCTTATAGGCTGTTGAAAAGCGCGATGTCC
|-----|
59083 GGGCTATATAATCCGTTTCCGAATGGCAGCGTGGATGTGTTTCG
59073
...
59214 TCTGTAGCGAAGAACGACGCAAAATTAAGCCGCTGATGCGATTAAGGTGACGCCGACA
|-----|
59274 TCCGTTAGGTGGCAAAAAGGAAAAACAACACCTGACGGTTTCTTTGAGCCGGAAGT
59264
59334 GCAACGACAGACGCTTCAGAGCGGTTTCCGCCGATCCGTCGAAAGCCACCATTTAGTGTG
|-----|
59345
59394 AAAGATATGACAATTAACGTTAAACGCGCTTGCGAGACAGGTAAGGTGACAGATCCCTGTGTG
|-----|
59423
59454 AGCGGAATGTCAGTTGCGCGCAGTGGCTGAAGTCACCGTTACTGAAGCGGCGCTGCA
|-----|

59514 GGGTAACGAGGTCAATACATGTTTCTGAAAAACAGAACAAATTTGAAT
59505
...
59662 AACTGCAGGTAAGCGTGGAGAGATCTGTCAGAACCGGCGGCTTCTGTGTGGCATGTCC
|-----|
59722 CTGTGGCATTAACCATTCACAGAAAAAGCAGTCAACGTTCAATGATG
59712
...
60214 GGCCTGATGAGGACAAAAATGCTGATGACAGAAAGCGGACGACTTGGCCGAGGCGTCCGT
|-----|
60274 TCGGTGGGAGCGAGGGCGCGATATTTTATCGTCTGGCGGATGTGGCGATGTCATGTGG
|-----|
60334 ATGATGCCGCATTAATGATGGCTT
60302
...
60414 TGACCTGAGTCTGATGCTGTCCGTTTGGATGACGAGATGACCGGGTAAGCGTCATTTT
|-----|
60474 TTCAAGTCTGATPACCGAGCTCAGAAAAACCGCAGTGTCTTGAACAGAGGCTGAGCGG
|-----|
60490
60534 CCAGCGCTGGCTGCACAAAAAGCCGGATTTCCGTCGGGCAGTATTAAGCGGCATGCG
|-----|
60537
60594 AACCTGCCCGCACAGTTTACGATATCGCCACGACGCTTCCGGTGGTCAGAAATCCCTG
|-----|
60654 GCTGATCCTGCTGCACAAGG
60619
...
60724 CTTGCCGCTGCATCACCTGCCGATGTGTCGGGGTCACCTCGCTGGCGGTGCGACAGT
|-----|
60784 GCGCTGTGTACGCTGTGTACGAGGAGATTTCCACGCTTTTCAGCGTTTAATTAACCCCTG
|-----|
60844 GTTCTTCCGTTATCACTCCGAGTGACTGCCATGCTGCATGCTGACGCTTCAAGAGCC
|-----|
60904 GGGCAGGACGAGGCTGACGTTTAACCAAGCGAGAGAGTCACTGCGACCCCTGTGTAAT
|-----|
60964 GCCCGTGTGCGTG
60921
...
61113 CCAAGTCCGTAACGTGACGCGAGAGAGATGCGTATGTTGCGAGCTGACAGCGTTCCGG
|-----|
61163
61173 TGATGAGGCGCGGCTTTACAGCGGCGGACAGATATCGCCACGAAAGGCTTTGATGACGA
|-----|
```

```
...
54883 GATGTCGCAGATTGAATATACGGCGCTTCTCTGATGAGCTTAACCGCGCATCAT |-----|
54943 GATCGGAGTGGTACTGGGAGCAACCCGGGCTCGCTCGTCCGCTGGAGTGAAGTTA |-----| |-----|
54971 54980 54974
55003 TGAACGACGCTGTCTCTGATGTATGTGAACGGGCTTCCGCTTCGTGAGAGTCTACAGT |-----| |-----|
55010 55035
55063 ACAAGGGGGGGTGCAGAGATTGGCGG
...
55223 TAAGATCACATTTGCGGCAATATGTTTCTTATCAGTACCGTCCGAAGTGGCGCT |-----|
55283 GGGGATGCGGGAGACCGCAGCAAAAAGCTTTGTCGATGAGGTGAG 55273
...
55369 TCGACGTGGAGAGAAAAGCAGCTTCACGGAATTATCCGTGAAGTGTGGCGCTCATG |-----|
55419
55429 CGTTTAAGGGGAATCTTTGTGCAGCGCGTCTGGATACGGAAC |-----|
...
55592 TGCCGTTGCGTACCATATCTGTGAGATGATTTCCGTCGTCGAGAGTGGAGATGGGA |-----|
55642
55652 ACGGATCCCGGTGAACCTTCCACCGGCGTCGCCCATGCTGCAATATTTTCAGACCGGT |-----|
55678
55712 GGAGGAGGGGAGACCCGCTGGGGCCACAGTTTACAGCCTCATGAGCGCTGAAGAT |-----|
55749
55772 GCTGATTCCTTCAGAGCAACACAGCTTCAGTCGGCATTTGTGAAGCCATGTATGCAGC |-----|
55829
55832 |-----| |-----| |-----|
55834 GACGATTGAAAGTGACCTTGATACCGAAAGCCTTTGAATATATTCGCGGCTGCGCGCA 55875
55832 55832
55892 GGGCAGAGGATATCCGCTTATTAAATTTCTGGAGAAGTTTTCACAGCTGGTATGACAC
55952 |-----| |-----| |-----|
55961 GAATTAAGTACGCTGGGTGTGTCAAAATTCGCAACCTTTTCCCGGGATGATCTGA
56012 ACTACAGACTGGCGAGATTTCAGCAATGATTTTCGGCGCTTGAACGCGCTGCTGCG |-----| |-----|
56048 56060
56072 GTATATGCGCGCGCGCTTGGCGTTTCCCTACGAACAGTTGTCGCGTATTAAGTGAAGT |-----| |-----|
56076 56091
```

```
56132 CAGTTATTCAAGTGCAGAGGCGCTTGCCAAATGATCTGGCGCTATTTTATGCGCGGCG |-----|
56192 AAAATTTATTGCGGCGCGGCTGGCCAGCAGATG 56170
...
56376 GAATGGTGTATGCCATTGAGGCGCGGACTGAGCAGCTATGAGAAAGACCGCGCTGATG |-----|
56436 GCGGAGATTATTCAGACATTTTCCCGCAGAGGTCAAGGAATCTG 56426
...
56730 TGATGCTCCGGGAATGCTGCTGAAACGAGCAGTACATGGCCGGAGGTAAACGACCGGC |-----|
56780
56790 CCGTGTTTACCGGGTGTGAACGGTATTTGCTTACTGCGCGTGACC
...
56852 GCGTGGGGGTATGCGGCCATTTTCCGGAATGACAGGCTATGACGCGCATTTGCGCTGTC |-----|
56902
56912 TTCAGCAGGCAATGGCGGATAGCCAGTGGCGGGGCTACTGCTGGA
...
57034 CCGGTCTGGGCACTGTGCATGACACGGGCTGTCTGTGACGCCATGCTGCTGCGCTGGGC |-----|
57084
57094 TGTCCCGCAGCGCTGGTTACCCAGACATCCGTTATCGGCTCCATTG
...
57406 CTGGCGGATGAATTAGTCATGCGTGGATGCCATCACTGTGANTGCCACGGCGCTGAAC |-----|
57456
57466 AGTAATGTCAAGAGGAGCATATGCGCCGCAATTAAGTCAAGGAAG
...
57563 CGAAAGACGTGAACAGCTTGCACAGATGCTGGCAGACAAACAGGGCATGAGCGTTGAAC |-----|
57613
57623 AGCCCGGGGATTTCTGGCGCGCGGCGGACCGCAGACGCGGTGGCANTCCACGAGAGTG |-----|
57630
57683 AAG
...
57688 GATGCAATTATGGCGTGTGAAGAACGAACGGTGTGAACAATGGCGGCAACGCTGGCG |-----|
57738 57747
57748 GCGATGCCGAGATGACGGTGAAAAAGCCCGCGGATCTGGGCTGTTCAACCGC |-----|
57752
...
57863 CTGAGGCGCAGGCTGAACAGCTGCTGCTGCTCCCGGAATGACTGTGTGAGACGCGCGCG |-----|
57913
```



```
51758 GGCCTGCTCTCCCGCTGGAAGAGCGCTTCATGGCTATGTCACTGACGCGCAACTCC |-----|
51780
51818 ATGCCGATGATACCCCGG
...
51868 AAGACCGGGCGGTTGTGGCGGTATGTTCGTGATGACCGCAATGACGGGTACAGCGCTTGCA |-----|
51918
51928 CCTGCAGTGTGGTTGCTTACAGCCCGGACAGAAAAAGCATCCATC
...
52167 CGCCATAGAGCGGATATAGGGGAATGCGGACAGCAGCGGCTTGCTGAACGTCACGCG |-----|
52217
52227 AAAAAAGAACCGTTGTTGAATCCCTGGAAGCTGTTGCGTCAAAAAAGATGAACACCT |-----|
52287 GTGCGACACTCAGAGTTGGCGAAGCGCTTGCGGTACGCACTTAACAGTGGCGCGCACT |-----|
52311
52347 GACGTACTATGCGAACGATGGCTGGGTGGAATCGAACAAACACATCGCTGAATAATGCCCT |-----|
52360
52407 GCGGGCGGTCACTCGGGTGTGTAATAACTTCTGTTCTTCGGCTCTGATCATGTGTTGA |-----|
52467 GCGGGAGCGCTACTGTACAGCCTGATCGGAGCGTGAACACTGATGACGTGATCCAGA |-----|
52473
52527 AAGTACCTTGGCCATGTGCTTGGCGTCAATGACAGACTGGCGGTCAACCGGTCAGCGA |-----|
52549
52587 ACTGCTTCCGTGGCGCAT
...
52685 AACTGATTGACACCGGGTCTGTATTTTCCCTCAGTAAGCTGGCAATGGCGTTATCC |-----|
52735
52745 AGTACGCTTGATTCATCCCTTTTATCCATGCAAGCAGTTTCTGATTTAAACCGCGC |-----|
52776
52805 CATCTTGACCATCTGAAAAACCTTATT
...
52873 ACCGATTTTACTTGATGATATATACAGCAACAAGTAATATATGATGAGCGCTGCA |-----|
52923
52933 CAACGCAATTACAGCCCGCTGATAAGGCATTGCGGTGACGCCAGTTGAATGGGCTGATC |-----|
52970
52993 AAAATTATTATCTGCTTAAGAATCTTCATATGTGTGAGGGGGAATGGAATAAGGCTGCAT |-----|
53042
53053 TCCAGATCGCCATCATGAACAGTATGGGAATGATCAGATCCGACCTGTTAATCTGATTA |-----|
53058
53113 A
...
53196 TTTCAGCCCGAGGATTTGCGCGCTGAGATTTTATGAAGTCTACGTGAGCGGACGATT |-----|
53246
53256 CGGAACGTGCCATGCGCTGAAGAACCTTTCCTCATGTGGTGTGTAACATCGTGACAAAT |-----|
53316 ACTCTACGCTGAAGACGCTTTTCATCGGCGCTTTCTGTGTGCTGGCGGCGCTGCC |-----|
53329
53376 GCCAAAAACTACCGTGAAAAAATCCGTGACGTGAGTCTGCTATGACGA |-----|
...
53475 GGGGATTAAGCGTATTGAGGGGTGCGGTGTGGCCAAATCCATTGCGGCTGACGCTTAAA |-----|
53525
53535 ATCAAAAGGCACTGCCAGATCGAAAAAGCCGCTAACAAGTGGCGC
...
53590 GTTTTATGTGCCCTGCCCGCACTGTGGGGAAGCGCCAGTATCTGAATTTGGCGATGAGT |-----|
53640
53650 CCACGCTTTTGGGCTTAATGGGAGAAAGACAGCCCTGAAGTGT
...
53754 AACGGCGGTGATCTGTGAATAACAGCGGCATGTGGACCCGTGACGGTCTGACTTTTTC |-----|
53804
53814 AGCGCCCGGGGATGAATAATCCGCGCGGCTCCATCATGTTCATATCTGAGC |-----|
53814
54125 ATGGGCTCGGGAGAGGAAGCCTTCTGTGATTAATCATATTATGAGGCGCTCGA |-----|
54175
54185 TGAGGAAGAGACGCGTGTACGTGTGATGCGGCGATCAAAAAAATAACCGCATGCGGA |-----|
54194
54245 TGGACCGAATAAGACTATTTCCTG
...
54597 GAAAAAGGAAGATGCGTCTGCTGTGGGATAACAACAAAAAGCGGCTAAGCAAGCGCTGAC |-----|
54647
54657 TGCTGTGTATGCTTACGCGGCATTACGTGTGTCGCAACGCTGCGACCTGTATCTG |-----|
54697
54717 GCTGTACTGGCAAAATCCCGGAAGAGACAGACCGCGCAACCTGAAAGAACTGCA |-----|
54746
54777 GCGAAGCTGTCCGAGAGTGAATGTGTTACGTGCTGAACCTGCAGGCGCTGCGTCAAGC |-----|
54808
54837 GCGCCTTGATTTATTACCGGTAAACGGGTGTCTGTCC |-----|
```

```
48494 CAGAGTTACTTCTTGTGATGGCAGATAATAAGCAAGCAACAGCAGCATTTCTTCAGACC
48554 CTTACGGTAGATCTCATATAAAGCGTGGATATTATTACCACACATTAAACCCGTGA
48614 TGCTTCAACGGCTGAAC
...
48894 TTTCATCAGCTCCTGTTATTGATTGCCCTGTGACCTTTTCTTACCTCATGACGTTTTTT
48954 CAGTTAGAAATATTCATTTTACAACCAAGTTCGTAATTGTTATTCAT
...
49536 TGTTCGCGGTTCTGTCTTCTGTTGACATGAGGTGCCCGTATTCAGTGTGCGATTT
49596 TGTATTGCTGAAGTTGTTTTTACGTTAAGTGTGACAGATCAATT
...
49696 CACGTTGTGACATGTAGATGATATTATTATTCATTTTGGCGGTCCTTCCGGGATCCGA
49756 CAGGTTACGGGGCGGACCTCGCGGTTTTCGCTATTATTAAGAAATTTTCCGGATCCA
49816 TGTCCGCTT
...
49953 GTGAGAGTGGCTACGCAGACTGAAGTGGCCAGGCATTTAAGTCTGACCGATGCCAGCT
50013 TCGCAGATTGCAGAAATTCGCGGTCGCCCGGATATCGAATAAGCA
...
50069 ATCTGATGCTGTGGCGGATTTTACATATCGTATCTGAGAGAAAGTAAAGCATGTGC
50129 CTGATGGCGATAGCGAAGACGACTATGAGAGAAATTGCTTATTCGACAGTGGAACTGA
50189 CAGCAGAACAGGC
...
50393 CCTGTGAACCCGGGATATCATCTCAAAACTTGCTCTTGAAAAATGSCATTAAGCCCATTC
50453 TGTGTTCAAATGGCGACACAATGGCGGAGGAAAGCTGCTATTACCTTCTTCAGAGA
50513 GCCCCAGCTACTCTCTGTGACTCTGTGATGAGCTGCCGACAGCCAGAGATGCTGCGAG
50573 AGGACCCGGAACCTCAGTATCAGCTGTGAGGTAACTGTTCCGCGCACGGGAGCTCCGCT
```

```
50633 TCAATGSCATATGTCAGCGAAAGCTCCTGACTGCTGTGATACAGAACTGAAGATGAT
50693 CCGGTTACCTTCGGGACCAAAATTTGGCTGTGCGGATATACCGATATGAGAAATGG
50753 CTTCAACGGCTGCTGCGAAGTACAAAGCGGCTGAAGAGATCCATGTCCGCCA
50813 TGTTCATATTTCCGGGCGCGACGCGACTCAGGT
...
50858 GTGGTCACCGGTGACGAGCTGTGCTCCTGACCAAAAGGCTGAGGCGGCTTCCG
50918 CTGCGCGTCAGCCCGTGTATGCAAAAGTGTCTTTACGACAGCGCAGCTGCTGCT
50978 GGAAGGATGCACTGGGACAGCCTAAGCGGCTGTGACCTCCCTGACATGCTTAAT
51038 CTCTTATCTGCTGTGTCACAGAAATAAGCCGGTAAATATGAGGCTTATGACGATCT
51098 CTCTGACGACATCTTCTCTCTGAAGAGGCTTGGCCGACAGAGAGGCTTATCCAGC
51158 CCTGACAGAAAGCTGAGCAACCGGAGCGGCAAAATAGACATCTGACAGCGCAGCTGG
51218 ATTAACCTCCCGGATGAACTTGGCAGTGTTCGCAAAAAGTCTCCCGCTATATGCAC
51278 AATGGAAGCGGATCTGAACCGGCTTCAGAAAGAGCGGATAGCTGATGTAAGGCTGT
51338 ATGACCCGCGAGTACAGCGCTTGCCTGTCAGACCCGCAACCCGTAAGCCGTTCCGTAAAT
51398 CACTACCCCGTACGAAAGCGGCTTGTGCTGCGGCGGCTGCTGCCGAACTGCGGCG
51458 GTTCACTGAGCTATCTGGGAGAGATACCGCGAAGCTGAGTGTGATGCGTAAAGGCTT
51518 TCCGGTTATTCGGAGGATGAGGAAAAAATGCTCTACTCAGTGCATGCCATGCTGTC
51578 AGGCACCTGCACCTTGGCGCCATGAGAGGCGGATGCGCGGAGCCGAGCGGCTGCGGCC
51638 GGTGCTGACCTGGAAGTATGCAAGACACACCCCGCTGTATGCGCCAGTCAGAATATAGG
51698 GCGGCAAGGTGTGAGCTGAGGCGTTCACTGCTGTGCGGCTGGGTGATGATGCTGCC
```

45631 CGGTGTTCCGGGTAATGGGCCCGCTCGCTGTGCACCATCATCAGCGATAA 45683 |-----|
45691 AGGCTCTAGTATGTCTCACTGGCCCTATACGGAGCGACTAAGAAAGACCGGATTACTGGC |-----|
45751 ATCAACAGGAGTACAGCGACTCGTATGACAACGGGATGGCTGAGAGCATCAATGGTCT |-----|
45811 TTACAAGCGGA 45767
45853 TGCAGAGTGAACCTGGCCACTACGTAGCTGGTGGAGCTGTATACATCGAGATTGCT |-----|
45913 GGGAGGCTGGGCCATCTCTCCGGCAGAGCAAGAAAAAGCTTATTATGCTTCATCGG 45903
45973 AAACGATGATCTGGCAGCCTGAGTTCAAGATAAACAACCTCTCCAGAAACCCGGGG |-----|
45974
46067 CCCAACAAAGACAAACAGCAGCAGCACCCTGTTCAGACAGTTTACTTAACGACTTAT 46117 |-----|
46127 AATGACACAAAAAGCGACACTAAAGTCGCTTTTCTTATGTAACAGCAATAAGCT 46153 |-----|
46187 CTCAGATATTTTAAACATTTTGTGACCGCGCTTCCGGAGCTATCTGTTCTCTCTG 46181 |-----|
46247 CCCTTATATCGTCGGAAATACCCGCGCTTTCAAATCCCATTCACACTCAAAATGTAG 46255 |-----|
46307 TCTG 46447 |-----|
46397 CGCCACTTTTCTTTCTGTGTCATCACTTAGCTGTACCATCACTTTGGCCATCTTCAGC |-----|
46457 TGCTCTCGGTACAGTCTTTTCTCCAGACATTCTGCCACGTGA 46447 |-----|
46639 AATAATCACCTTATTTTATTTTGTGAGGAGCAATTACTGAACAAAAACGGTGA 46689 |-----|
46699 ATATACTCAAAACCAACAACTATTCTGCCATCAGTATCATGGC 46689 |-----|
46832 TGCCCCACTCCATTTACTGTAAATAAATATCAAAATATTTATCACAGAAATCGTTTGG 46882 |-----|
46892 GCCATGAAGTACGACACTATAAGTCCGGAAGTACTCTTTGTTAAATTTACCTTAAGT 46947 |-----|
46952 TACGAGTAACACCTCATTAACAAACATCACGGTATACACTGGGTACGGAT 46947 |-----|

46952
47106 TTATCCAGCCGGAGGTGCCGGTGGCTTCACGCACGGTACCGAGAGTGGCGGTGATC 47156 |-----|
47166 CGCAGCGCTTACGACACGAGCGGCAACATCAGCAGCAGAGTTTCAATTTTCAGCTCTCGCA 47170 |-----|
47226 TCGGCTAATTCCTCGATATCTGCGATCAAGTGCAGCAGCATCACGCTGGCGTATCTGC |-----|
47286 ATATCAGTAATTTGTCCGCTTCGCCAGCTCCA 47261
47486 CCACATTACGAATGCGAAAAAGCTGCCGCTCTGTGTTAAATCCAGACTGGCTGGCGTCCA 47536 |-----|
47546 CTTCAATGCCCTTTCATATATGACAACTCAGAAAGATCTTTTATA 47536 |-----|
47705 CTTAGGTTTACCAAGTTTTCAGGATTTGCTGTATTCACCCCGGTTGGCCAAAGCTATT 47755 |-----|
47765 TTCAGAAATATTTCTGTAATCTGTGATGTCACAGCAGCGCAGC 47755 |-----|
47850 TATTGAACGATACGGCAGTCTCTGCCACCGCTTAATCCACGAGCAATCGCCTCAC 47900 |-----|
47910 AGGCACCTTTTGCATGCGCTGCATTATTTGTTTATTAAGCTCCAGCGGAAGCACTTAC 47923 |-----|
47970 CGGGCCCAATGTTGTACGAGCAGAAATGACGCGAT 47948 |-----|
48009 CTTTCTGGGTTCCGTACAGCGCACCCGAGTGTTCACACCATGCGACGCGCTTGT 48059 |-----|
48069 CAGGTTGATGCAATTAACCCGGTGCATTTTCTCTTGACAGCTT 48059 |-----|
48219 CATCCAGAAACTGTGTCAGAGATTTCAGGCGCAGAAACACCTCAGCAATCAGCGCAGAA 48269 |-----|
48279 CGGCAGCCGACAGCCGCTATTTTATTTTTCGTCATGGGGATTAATGATTTCTAATCC 48269 |-----|
48339 CTTGATATGTAGTATATATATCAACACTCATGTTGGTGGCTCTCATTAACATATCCCTTGA 48373 |-----|
48399 GACGACAGAGATTACACAATGAAGCAT 48373 |-----|
48434 TGAACAGTAAGAAAGTTTGGCAGAGATTTTTCACAACTAATGACAGAAACGTTTAAT 48484 |-----|

```
42289 |-----| |-----|
AATCGCCAGTGCCTGGTTCAAAATCCAGCAAGGCCACCATCATATACCGCATTAAGTTC
42290 42298
42349 |-----|
ATCAGAAAAGAGCGCCAGCTTTCGAAGCTGTTGCGGAGAGTTGCGGTCGCCGAAGCGCG
42359
42409 TCCATT
...
42676 TCGTGTAATTCACCGGAGGACCCGACCAATGCAATGACATACCGCATC
|-----|
42736 TCCAGCCCTCTCCGAGGGGCTTCTTATGACAAAAAACCCG
|-----|
42974 TCCGCGCGGAGATTTCATGSCATTAAACATATGATGTCAGGGCGCTGCCG
|-----|
43034 TCAGACCTTGGGAAAAGCTGACACACAACGTGAAGAGGCTGGCAGC
|-----|
...
43281 AGCTGGCGCGCGCAGCAGGTAACCTCCGGGTGAGAGAGTTGTACGTATACGACATCA
|-----|
43341 TTCGGCGCACCCTGCTGCATGATGTGACAGATATGATAGCTG
43331
...
43485 CGGGATCTCTGCTGCTACCATGCTGTCTGTTGTTTGCATTCCACCGGCGCTGAGG
|-----|
43545 GGACATTCAGTGCAGACAGGGGGCCAGCAGATTCCGACGCTG
43535
...
43656 CGAAAAATGTGTTGCTGGCGGTGTGCTGATGACAGGAGATTGACATGACGCGCCCA
|-----|
43716 CCCACGACAGCAACCTGCGCTGTTTACAGCCATGCTGACACAGTT
43706
...
43866 CTACCCATAGACACCGGTGACGCGGGTATAAAAACAGGAGAGTGAAGCGCTTATTT
|-----|
43926 TTGTGCCCTTATGACAGAGGTAAAGGCGTCAATACCGCACAATAACGCGCGGCAAG
|-----|
43952
43986 ATCCGGATATTCCGGCATCAGG
...
44053 AAACGCGCCGACACATTTCAGTTTCATGGGCGCGCAGACATTAATCCGGAATCGCTGCG
|-----|
44113 AACCGTATTCTGAACGACGCGGCGCACCCTCCGCTTCATACGT
44103
```

```
...
44453 TTCGGGCTGGGATTTATCTGTATAGGATGCTCCCGTTCCATGATGTGTGGCATGAGC
|-----|
44513 GGTACCGGTAAATCCGTTCCGTGATGTGTAATTCACCTCAGACACATG
44503
...
44568 TGAATCTGATGCATCACAAGAAAGCCGGAACAGACGAGTGGGAGTTGCGCATTTACG
|-----|
44628 GTAACGACTGGCAGAGCGCTGAGAGTGTGTTCACCGCGGACAGTGCAAGTTACTCCGA
|-----|
44688 AACTGAATG
44641
...
44708 CCGCATTCAGTTATAAAGACAGTCTGACACTGGAGCTGAATGCCGTGACGTGACG
|-----|
44768 GATGTACAAAAAATGCAGCGTATGGCGTTGTGATAGAAGTCTGCTGAGATAAT
|-----|
44828 GCACCGCAGCATTAATAAAGAGAGCCAGCACTGACCTGAAGAGACGCTGGCTAA
|-----|
44888 AGCCTTATATGTTTGTAGACCTTATTTTCACAGACAGCA
44856 44875
...
45009 GTAAACAGATGACTAAAAATACCTGCTTTTCCCGGAAAGTCCGTACGCGGCGATTGCT
|-----|
45069 ATGTTCTGAAAGTCAGGATGAATATGACTCAGACAGTGGCGGCA
45059
...
45166 GGGTTCGCCAGCATGAGCGGATTAACGGGGCGGTGATGTGGCTCACAGCGCTGAG
|-----|
45226 GTCAGCGTCTGAAGAGCTGGAACGTGAATCTGAACTGGCGCAGTAACGATATCC
|-----|
45229 45276
...
45286 TTCGCGCAGGCTTCCGCTTATTTTGGGAAGCGGAGTTGACCGCCT
|-----|
...
45391 ATGACGAGACTGATATTGGCCCGTCAAGCATTAACCATTTGTCACAAACAGCAGCATCA
|-----|
45451 TCCGATAAACGAGTGCCTGCGCAGCAGCACTGGCTGAAGAGAGATACACG
|-----|
45511 CGTATACATGAATAATCATCAGGTGTACGGTGTGCGTAAGTCTGCGTCAAGTTGTACG
|-----|
45571 GGAAGGATCAGGTTGCCAGATGTACAGTGCACGTCTCATGCGGTTATGAGACTTGC
45482 45555
```

```
...
39187 GATGGGTATAGATTTGTTTGGCTTACCCGCGACGCGTGAACACCTACTGGCGACGTGC          |-----|
                                           39237
                                           39240
39247 TGGCAGCACATATTTTGTATCAAAAAGCCGCTGAGCGGTATGCGCGGCTGAGCATAT          |-----|
                                           39279
                                           39285
                                           39288
39307 TGTTCGCCAGCAGCGGCTGAATAATTAGCCTGTCCGGACGGCTGGCAATAAAAATTTATTGC
39367 AGAACCAACCGGATTAAGCGCCGCGTGACCTGGACACATTTCTGAAGACCGCTGGATGC          |-----|
                                           39381
39427 GCTGACGCAT
...
39495 CTGTGCCAGGTGACGGCTGGGCGTGAAGATTTCGAATCAGAGTGGTATGCGATGGCG          |-----|
                                           39545
39555 CGTGATATCCAGATGTTCTTGACCGATGGGGGGCGCATGGCGCAATTAATCATGAAGAT          |-----|
                                           39577
39615 GTAACATGCCCCGCGATA
...
39646 AAGGATTAAATCCCGACTAAAGTGAATACAGCTCCTAAGTGTCTGATGATGAGCGCCATGA          |-----|
                                           39696
39706 TAAATTGTGGTTGTATGGCAGATTAACAAGAAATATCAGTATTT
...
40160 CTGTGCTTAAATGATATTATTGATTGGCGGTATTATTTTGTGCTACAAATCGTTTTTT          |-----|
                                           40210
40220 TTATTTATTTGAAAGTTCACTGGAATATTAAGCGCTGGCTAAATCTATATATGAAGTTG          |-----|
                                           40249
40280 GCATAGTTTTATCAGCCATAGCAAT
...
40357 ACTGTGGCCCTCAAAAGTACTCCCTATACAATGATTTTAATTTTGAATAAGCGCTGTG          |-----|
                                           40407
40417 CCATTTTGTGTAGTCGTGTACATCTCTTTGTTTAAAAAAGAAATCAAGAAATCAGAAAGTG
40477 TCTTTGTGGAAGTGGGGGCTGTGTGCTCTGTCTCTTATACCTGTTTTTGGCTGTT          |-----|
                                           40500
40537 TTTTGTGTGATTAAGC
...
40621 GCAGAGCAATCATTTATTTAGAGGTAATTTCAATCAGCGCTATCAGAAGTAAAGTCTCCA          |-----|
```

```
40681 TTGTTGCATTAAATGTGGCGCTTGTGTTGGTTTTATACACT          40671
...
40752 ATTTGCCACGTTATNCGTGTGTGTGTATGATTTGTCAATGAAATGTAAGTGGAGCTTTGTG          |-----|
                                           40802
40812 GGTTCGCAACCTTTTCCATTTTGGTTGATCTGTGTACTTGTTA
...
41181 TGAACAACAGCATATATGACAGAAACCGCAGAGTTGTGTTGATGAATTAAGCGTCACCG          |-----|
                                           41231
41241 AACCGGCAGCGCTCGGGAGATTGCGAGAAATACGTACCTTTCAACGCGAGCGCTGCCAGT          |-----|
                                           41245
                                           41289
41301 TAAATCTGACCCAGCTGTTATCGGGGCTGTGCGACTATCATGTTCCGTTGTTACAGACG          |-----|
                                           41357
41361 CCTTCAGTCTGAAAGCTTTTTTATTTGTGTAAATGGCGCGCTGTGGTG          -|
                                           41362
...
41548 CAATTAATATTTTCTTGAAGAAGATAGTTGTACATTAAAGCCCGCTCAGTCTCGG          |-----|
                                           41598
41608 CCACCAAAATGCATCAGGCGCTTTTCTTATGCAATTAATGCAAAATTTTGCATA          |-----|
                                           41624
41668 TAAACCATCCCA
...
41683 TAACTCAATATCTGAAATTCAGCATTAAGCTGACACAGAAGCTCAATCTGACTGATCTCGC          |-----|
                                           41733
41743 TTATCTGTCACTGGAACCCACACACCCTGGCTGGCTGAGGCTTT
...
41799 CAGGATTTTCCGAGATACCACGAGATATGTTTCTTATGATGCAATCCGTGCTGTTT          |-----|
                                           41849
41859 GAATCAGATGAACGATATTTATTTTAAAGTATGTTTATCCAGTATTAAGGCCAAA
41919 GTCAGGCGCATTAATTCAGACAGATATTCAGCAAGCTGGACGCTTGGTGGAATGGCTTA          |-----|
                                           41957
41979 CGCCTGCATATGAGTGAGAGTTGTTTAAAGATAAACTGAATAATGAAGGACAAATGTTTC
42039 AGTCGCTTGTGCGAAGAGCGGATGCCGTGCTGTCAATATGTTATGTTCCCG          |-----|
                                           42040
...
42229 CATCTGATTAATTTATTTTTCACCCAGCCCTTTAGCTCAGTGTGAGAGCGCAATCAT          |-----|
                                           42279
```

```
35182 GAAACGCAAGCAGATGAGTATTTTTCGCTCCCTTCGACGGGCAAAACCTGGGGCTCCGC |-----|
35242 AGGGCGAAAAGTGATGTTTCAGAAAGTGGAGCGAGTCTGCCGCCGC 35232
...
35323 CAGATTACTGATTTCTCCCGCCGTTTGTAFCGGAAGAATGATTGCCGAGGCGCTTATG |-----|
35383 GCAAAAGTATTTACACAGAGAAGCGGGAATAATTAAGGAGGTTGTTGAGCTAGTA 35373
35443 CGCGGAGTGGCGCGGACGTTACGGCAACTGGAAGCCAAGACAGGTGGCAGACAGTAT |-----|
35503 CTGATGAGCGTCTTTCGACAGAGAGCTGGTTGCCAGTGGCGATGTAATCAACTCTGGTTAC |-----|
35563 GGGTTATTTCCCATCTGACAGAGCTCGTAAGAGCTGGCAAAACGCCGCAAAAACCTATCG 35539
35623 AGGGCAAAACTGAAGAACCAGGTTGGTTGATCC 35602
...
36619 AGAGTTATCAACAATAGCATTTCACCAATCAGAGATTAATCGTAAGCAGCGGATGAAT |-----|
36679 GCTAAGGTTATTCAGATTTCATTAACCAAGTGAATCACTACACTGGT 36669
...
37002 ATCTACCTCCATTATGATTTTAAGACTGATATATGTTATCAAGCAAGGTACATGTTTAA |-----|
37062 TTTTACCTGTCATPACATTTGATATTTGATATGCTTTACATGCTTCTGCTTAGAATTTGA |-----|
37122 AGGAGGTAACATATCATGTCAATGAATGAATCAACCCAC 37092 37101
...
37240 GAGGTAATATAGCATCAAGCAAAATCTCTTTTCCGTTGTTGCTTATGAGACCTCTGTC |-----|
37300 TTTCTGACCGATTTTCATATTTGGCGAGGTAATGGAAAGTTAAGTAGA 37290
...
37386 CACCAAGCGCAGATAGAGAAAAGCCCGAGTTAAACATTACCGCTCTCAAGCCCTAAC |-----|
37446 ATTAATCTGAGGCTCAATCTATGAACGGCAATCTAGGTTAGCCTCTTACGTGCCGAAG 37436
37506 GCAAGGAGAGCAGGCTATGAGAGAGCAAAAGGAGATGTTAATGCGCCCTGATGCTATCT |-----|
37566 GTTTAACCGTTCATAGTGACGCACTGGTAAAGAGAAAGACCTCTGCGAGGTACGAGTCC 37537 37547 37556
37626 GAACGGCCAGACGAGAGGTGCGTGTCTTCAACAGTTTCAAGACCTGAGAGGTAAAGACCT |-----|
37643
37686 GGGCGGGGAGAAA 37643
...
38124 ATGTGGGGGATATTACTTCGACCTGTTCTGTGCTACCGAAGTCTGCTGTGTTATGTAAGC |-----|
38184 CAGGCGGTGAATCAATGTCAGCATTCGATTAACGGCAGAAATACTGCT 38174
...
38277 CTGCCGTTTCAACAGCCCGCTGGCGAAGATAAATCACTACTGTCATTTTTCAGGATGAGC |-----|
38337 GGGTATCCGCTGCTGCGGGTGGCGCTGTGTCACACTGTCGACTGGTTATTAAGTACGTCGA |-----|
38397 AATCTGCGCAGTGGCCTCATGATGATTATCATCAACAGCAAACTGTCATCATCTTACG 38448
38457 GTACCGCGCGGATGGTGTGTGTGCTGGCACTGCCAACAACAGCTGCGC 38448
...
38514 CCGAATCACTTGAGCAACTTGCTCAACAGAAATCGGCCGCTGAGTATGATGACGTCATCC |-----|
38574 GCCACGCAATGAATGCGATACAGAGACGGGAATTAATGCTGGCTGAATTAATCTGGTGG |-----|
38634 CAGTCTGCATTCAGGTGCTGAGCCGATTTACCTGAGGAGTATCGCGTCTCTCTGGGAT |-----|
38694 TACCGGGGAAAAAATCCGCTCCGTATATCCGTGAAGACGACATCATACCGGGAGAACAGA |-----|
38754 CCGCCACCACATTACTGAGACGCGCAGAAA 38729
...
38839 GACGGTGTCAGCATTTGCCGTTGATTCGGAGTCTCCGAATCCCTTCATGAAGACGCTAA |-----|
38899 ACCTCGCGCGCTGGGTAATGAGAAATPACACAGCGCTGGGTAAGACACAGCGGTGGGTG |-----|
38959 TTGTGTAAACCAAGCGGAGCATCTCTCATCTGATTTGGTCAATGTCAGGGTGAATGGG |-----|
39019 AACAAATCCCA 38975
...
39041 ACCTTACCGCTGTGTGCGGAGCATCACACAGAGCTTCATCGGATCCGCTGGCGCTTGAA |-----|
39101 GAAAGCATGTTCCAGTTGATTAATTTTTCGTTTCTTGATC 39091
```

```
31871 GTAGACGAGGAGGCTGCTGTAAGATATTGATGATTTGCTGCATCGCGGATCGTT |-----|
31921 31925
31931 CTGTCGATCAGATGATCGTGTGCCCCGAGGAAACACTGTGCTGTGATTAATG
31991 GGTATATAATGAGTAAATGACAAAAAATTGACACTGCTTTAGGAAAAACAGTCCGA |-----|
32051 CATCTGGCGTGGCGGTTTATCTGATGTGAGCTGSCAGAGTGTGTTACATTAAGCTAA 32028
32111 TGGCGTGTGCGGCTTTGTCACCGACGAGAAACGCGCAAGAAACAGATAAAGCTGGC |-----|
32171 TGATGCA 32122
...
32405 CGCACTTCATTAATTAGAGAGACACTATGGCGCATACAGATGCGAAGCGGTGATTA |-----|
32455
32465 CCTGAAGGTTATGATGATGATGCACTCTCTTTTGAAGTTCGCTAA
...
32971 GATGACAGATGCTGTGGGTCATGAATAATGCGACGCTTGGGGAACGCTTGAGAG |-----|
33021
33031 CCTGGCGATTAACCGTGTGTGCAAGTGTGAGGATGATGATCATATTAACGCT |-----|
33034 33085
33091 TTTCTGTACGTGTTTACGACAGATATAAAGGATTTCCATCTGATGT
...
33302 AGACCGTCATCTGTGTACGCCAGCAGCAGCAGACCGGAAAGCGGATTCGGGCGCTGTC |-----|
33352
33362 AGCAGTGTATGACCGCAGGGTCGAGCAATCAACCAACTGCAGCA
...
33576 GGCAGATTGCGCGAGGGGGTGCACCAAGACCATGCTGTGATGGTGTAGGCGTTACGC |-----|
33626
33636 GAGATACCTGGCGCGCGACCGGAAAGCGGAGACGAAATTTGTTACGAGCGACGACTT |-----|
33648 33684 33687 33690
33696 TTTTGGGCGCGACCGGAATTTTGAAMACCCCTGTGTCTCCGGTAAGCGGCACGAACAA
33756 CCAGCGTTGTGTGAATCATATTCTGAACGGATACGAAATTCGCCGGGATTCAGG
33758
...
33832 AACATTGCGGACCGCGGGGTTCTTGAACGTATTTCGAGACTTGCACCAACAGCGCTCGTT |-----|
33882
```

```
33892 CCACCGTACCGGACGCTGGAGAGATGCGGGGAATGSCAACTTCTG
...
33954 CGAGAGATTAAACCGCCAGAAATCACCAGTTGGGGGTGGAATAATCTGAAATCGTGGG |-----|
34004
34014 CATCCAGCTCTGACACAGCAAAATGCTCGTTTGCGAATATCAGGTGCAGAACGAGGCA |-----|
34063
34074 AAAATACCGCGCTGAGCCAGGCCAAATTCATAGCTGACGAATGAT
...
34132 AATTTTGTCTCAGGGTAAAAACCGGCACCGGAAAAATCACCTTGACGCCGCGATGGGT |-----|
34182
34192 AACCGCTGATGGCGAAGGGGCGCAGCGTGAATTAATGTCACCGCTGTCTGATGTATGAGC |-----|
34225
34252 GTGTTGCATGACAGCTACGACACGCGCA
...
34292 AATTTTACAGAGCTTTGCACTGTTGATTTGCTGTGCTCTGAGTGAATAATGCGCTTACGC |-----|
34342
34352 GGGAGACGAAAAAGAGCAGGTGATTAACACAGATTAATTAATGCGCGGACAGATCAC |-----|
34394
34412 TGTGCACTGTGGGATGTTAAACAACTGAATCATGCC
...
34465 CTTGTGAGAGATTATGACCGCATGACATGAACGCTGTCATGAGTACGTTTAC |-----|
34515
34525 TGGATAGCTGGCGTCCAAATGTCAGCAATATGAGGGTGTGAATATTTTGTCCGAG |-----|
34535
34585 GAAATT
...
34664 GAATCAGTCTGGAAGAGGCTGTCAACGAGCTGTGGAACTCAAAATAAATGCGTCTTG |-----|
34714 34717
34724 ATAAACTGTGTACACCTGTGTTTCTGGCTGGCAAGTGAAATCCCGGT
...
34964 GCGGCTGGACGCGTTAATCAGAACGTAATAATTCGTTACTGCATGCCGAGCATTAAT |-----|
35014
35024 TACCAGCAGACCGGAAAGCGCGCTGGTAACGAAAGTGAATGATTA
...
35122 CAGGAAGAAATTAACAGAAACTGTGGCGACATTTGTGACCGCTTGCCATGCTTAC |-----|
35172
```

```
28582   ACGAGTAAATCTATACCGCCAAATGTACACAGAAATATTGATTAAGAGAGCGGTTGCAG      |-----|
28642   CGGGGTGTGACTTAAGCGTATGTCAACCTGCACAAACCGGTGTC      28632
...
28726   CGGCATTTCGGCGATTGCCACAAACGGAAAGCGCGTGCATTTAAATTAAAGGACGGCT      |-----|
28786   CTAAGGAGAGAGACTTCGCGCGTGCACCTTTCGTATTAGCCCTGAC      28776
...
28931   TTCAGCCCTTAATGAGAGATAAATGCCAAACAACCATGCCCTGCCCTGTACGCTCTTT      |-----|
28991   CTGGCTTTTATAGAAAGGCGGTTATTACATTTTAACTGTCCAGA      28981
...
29072   TCATTACAAGCCAAATCAGTATCAGTCTGATTTACTGAATAAAGATTAAAGCGCCGC      |-----|
29132   GAGATTAAAGGCGTAGTGGCCCATTAATAATATTGCAAGACCAC      29122
...
29358   ACGAGTTCGGTAATCCTGTGTCATCTGAATAACTTTCATAGAAAGAGGATTTGCC      |-----|
29418   GCAAGAGATACAGTCTCGCTCTGCAATTTTTTGTAGTTTCCA      29408
...
29602   CATATTATCAATGTATTCTTCTGTCGATTCATTAACAATTCGTTGACAGCGTTG      |-----|
29662   AAGTCAATGCTGCAATTTCCGTGATTTCTTTAGTAATCTTTAAATTCATCTTTAA      29652
29722   CTCGGTTAGTCGTTAGATTTTGGCCGAACTGAAAGCGCTGTTTAACCTCAGAACTG      |-----|
29782   AGAGCTTCCTCGCTTCGGCAAGGCTTC      29755
...
29833   TCATGATTTGTGCAAAATCATATCACTCACCTGATTTCTTCCAGCCAGCGAGGCG      |-----|
29893   ACCATTTCGGTTTAAAGCTTTGCTTTGGTATACGTATCGCGGTAAATGTACCGTC      |-----|
29953   CTGTTTGGGAAA      29909
...
29982   ATTGCTGTTGCCAAGATGATAGTATCCATGTTGACCTCATTTCCCTTAAAGCGCGGT      |-----|
30042   AGCGAACAATAAAGCTGTCATAGTTATTAAGTTGAACCTGCC      30032
...
30107   GGCTGGCTACTTAACCCCTGACCACGTGTTGTAACGAAATATTGCCCGGCTTCTGT      |-----|
30167   GGGGGGGGTGGTGGTTGGTGTATGTATCTAATAATAAAAACTG      30157
...
30260   TCGAAAGAGGAGATATGAAATTTGTTCACTCAGATAGAAAGGAAATTGTGCGGAGA      |-----|
30320   ATGTGTCAGCTTACAATCTCAATGCTTGAAGCAGCTACTGTAC      30310
...
30462   GATGTTTCAATGGCTATTTTATGCTACGTGCTGCCGGTGGTTCTCTTGGGATGTAA      |-----|
30522   GTTATGATCTGGCGCAATCATGCGAGCAGACGGTAGCCATATAA      30512
...
30627   CAGCCATTAACCTGTCATGTGAATTCGGGGCGAGTGTCTGCAAGGCTTGACCTTACT      |-----|
30687   TACTGCGCAACCATTCGTTGTAACGTGTTTCTGTTGGTAGTTTC      30677
...
31136   CATCAGTTTCTGAAGAAGCTTGTGATTTCTGTACATAGATTTCTAGAGGCTCATC      |-----|
31196   ATTATTCAATTTTACAGTAATTAATCTGTTAAAGCATCAATAAAA      31186
...
31311   AGATACACTCGGGAATAATTAGCGAATAATACACTCAGGCTGAATAATGTCGTTATAT      |-----|
31371   GGGGTCGCTCAACACAGACTGTTGGCAATGTTTACCTTTGGGCTTCCGCCAAGCAGGT      |-----|
31431   AATTCCGTTATGCCAATAATGAAGTGGGAAGTTAC      31411
...
31541   CAAATGCGCGGAGTTGATTCATGAAATCAACATGAACACATCCGATGGGATGAAT      |-----|
31601   GCCTGGGCGCATCCGAGCGGCAAAAGTACCGGCTGCGAATAATTACAAAGGCTATTTC      31591
31661   GAGCTGGGAATGACGTTCCCGGAACGTATGAGCAGCAGCATCCGAAGCCATGCTCGC      |-----|
31721   AATACTCAGAAAATTTTCCGCTGGTGGAGAAAAGACACTCTGATCGGTTAAAAAATT      31691
31781   CAGGCTGTGTTACAGCGATCGAAAAAGGATGCCGCTCCGCTGGTGGCCCGAATGCC      |-----|
31841   AGCCACAGTTCGCGTTATTTTC      31807
...
|-----|
```



```

25371 TGTGTCACGTGCTTTACAGCATGACCGCGCCACTATAGTGTGGGGGGGATATGG
25421
25431 GTTGAACATCAAAAAATCAGCTGACTCGGAGCAGTTTGGCAGATGGCTGAACGTAT
25491 TTCGCTGGTAAAGTGCATGTCAGAGGGCGGTATGGCGTTTCACCTTCGAATGGAATATAGTC
25517 25517 25525 25548
25551 --|
GTTCTTGCGTCGAAATCCATTCACAGTCCAGAGCAACCATGCTTCTTACATGCGGATCA
25553
25611 TTTGTTAACGCCACAGGGGAGGAGTCCCTGCTTCATATGAACGGCGTAAATCAGC
25618
25671 TTG
...
26019 TGATACCGTTCACCTTTAATTAAAGGANTTGCCTGTTGCGCAGTTTATATATGCGTGGCG
26069
26079 TGATACTGATGCGTGGCAGCTGTTCTGTTGTCAGATGTTCAAGTGCCTTCCATTTCCCTCC
26087
26139 GGAT
...
26340 TGTCCGCTTCTGCGCTATTTTGTTCACGTTGGGCTGATTCCTGCTCAGCGTCCGGA
26390
26400 CTCTGATCCCTTTCACCCATTTGCGATCATTCGSGTGGCTTAATCCCTTCAACAAATTCA
26435
26460 CCACGTGATGACGACGAAGCAATTTATCGGCGTCCAGGCTGCTGATATATGGCTCCTCGATA
26486
26520 ATTTGTTTACTCTGTCAGCGG
...
26617 TTTTTCGGGCAAAATATTTCTTTTGTGAGATTTGCGTACCGCCGCGCAGCCGCTTGT
26667
26677 CCAGACGAGAAAGTTTGTTCGCCGCTATTTTCCCGGCCCTT
...
26851 GAGCATAGTAGAGCTGTGTCTTCTTCATGTTCAACACGCGCTTGTGTCGTCAGT
26901
26911 ATTTCATGATATATTTGTTTCAGTGCACACGGGATCTTGTGAATATGCAATCGAGCTTCGC
26962
26971 GAACAGGCGCAAAAAATAGTCTGACGCTCATAGCCAAAGTGTCTGGGGC
...
27031 ATGCTGAGATACGCTTCCAGTCTTTCGGGATCCCTTGTGATTAACCTCAATTTGTCGCCAGC
27081 27088

```

```

27091 --|
GATGATGGTACCCTCAATGTTTCTGATCCACATCACCGGGCCAGAGGCG
27093
...
27177 TGTGTTATATTCGGGCGAAATGGCAACAGTAGACAGCCCGTTTTTTTTCAGCAGCGTTTCA
27227
27237 GTGTCTGTTGTTGTTGACTCTGGGGGGGTGAGATCAACAACAGACGTATTTTCCGGTT
27297 TCTTCGCTTACCTTCGGGAGCTTTTTTTCAGATGCGCATCTGCTGAATTTTCAGGC
27362
27357 CATTTAGCGCCAGGCTCACATTTATGCTTAACCCAGCCGATGCGATGATTAAGTTCA
27362
27417 GGATACATGCGCTTACCTTCTGGCATTTTCAATCAACGCTTCAACGATATGTCGCCGAAT
27425 27450 27458
27477 GTTCCATGTCCTTCTGTAGTAATTCCTGTCGCTAA
...
27549 AATTTAACGACGACTGCTTCTGTACTTTCAGGGGACAGCTTCAAAATTTGACGTTTCATC
27599
27609 GGATCGAATTCGTGTTGCGACCGGTACAAAGAGCTGACTCTCAT
...
27758 AGTTCGCGGGAATGCTCCAGAAATTTGCCAGTGAATTTTGGCGTGGCGGCTCAATC
27808
27818 ACCGCTTCATATGCGAGTGGCCACCGCAACAGCGCTATCTTCTTTATGGCCTGTTCGTC
27848
27878 AGAATAGCAGCGCAAAATAAGATTT
...
28241 GGAATGATACTTTCATTTAACAGGTCTGTCGAGTTTCTGCTTAATGCGCTTCA
28291
28301 ATTAATGTTTGTCTCCAGTGATTAAGCCAAATTTGAAAGCAAGTTTCAGACGTCTGCA
28324
28361 TTACATGATTAACCTTTTCCA
...
28383 TCGGTCCTCAACAATGAATTTTGATGCGGTGCTGTCCTTCACAGTGCATTAAC
28433
28443 AGTTTAACATTTAAGCGCGGATACAGAGCATTTCTGTTACTCAGCAAAATGACCGCTTAAAC
28454
28503 GATTTAA
...

```

```
22171 GCTGGCTTTACCTTCACATACGGGCAACGGCCCCGGAAAAAGACCCAGGATGGCGACGAAG 22221
22231 TAATGATTTTTGCGGCCCTTTGGCTAACACTAAGCTACACTCATATATTTACTTTCTGCT
22291 CGATGCAAAAGAAATGGCGCTCATGTGTCCGGAGTTTTCATGAGATTTCACACGCTCCATT 22306
22351 AAGGATAAGA
...
22437 CACAGTACGAGCGAAAAAGACTGGGCTTAACCGGGTACGCCAAAAATCTTGCAGATGGCA 22487
22497 GCGTGAAGTGTGTGCTGCGGTGAAGAAGGCGAGTAAAAAATT
...
22716 GTGCGTAAATAAGCGGCTGTACCTTTTCTTGCCAAATTTATTGCCATCGCTTTTA 22766
22776 CCACATACGAATCGCGCGAGAGAATGTGAATTCAGATAGAAATCAGCAGCAAAAAACGA 22787
22836 CCACCTCCAGTGTCTGGCGCAGCGAAATCCCTTCGTCTGTGCAATCACCACTGCGCA 22853
22896 GCGGCTCAGTCCA
...
23150 ATATAGTTGTTTCACGCGCATGATGATGTTGCTGTTTCAACAAAAATACGCGCAGAG 23200
23210 GAGATCAAAATGAAGACCGCGCTGATCGCAGATGACAGAGCAGCGAGCTGCAGAAATA 23234
23270 TTACGACCATACCAATCAGCAGCACACACATACCCGCCATCAGCATACCGCCGAGAGAC 23327
23330 GACATATCTTTGGCGGTGTCAGCATATGACAGAGCAGAGAACACATACGCGCTT 23332
23390 CCGCCCAAGTGCATACGATTACGTACACCATTCGCGAGACAGATAGTATTCAGAAAT 23404
23450 GGTCGAGAGA
...
23508 GCTTATTCGCATTTTATAGGTCAAGAACATCAAAACATACATACCCACAGCGTCAGAA 23558
23568 TCAGACCCGAGATGCGAGCATCAGCACCGTACTGCGAGTTGCGGTAATCGCCGAAAAAG 23615
23628 CCAGCGTCAGGCTCAGCAGAAAAATAGTATTACGACGACCTTTATGGTGCTAAGCAG 23630
23630
```

```
...
23814 TTGATCTGACAGGTTTAAACAGCGGATTAATCAGTCAATTAAGCAATATAAAGCCCTGA 23864
23874 GAATTTGCGCAGGCAAAAGAAAAAGGGTTAGCATTTAGCTAACCC
...
23921 TTATCTTATCTGCGGAGAGCGCAGAGATTGAACTGTGAACCTTTTCGGGTCCCGCTT 23971
23981 TTCAGACCGAAGAAAAACAATAAATAATCAGTGTATACGTGAAA
...
24256 TAATCTGAGATCCCTTGGCTTTTATGTCGTAATAATTCACCCGAAAGGAAAGCCCACT 24306
24316 TTTTGCTGGCTGCGGCTTTAGCCTGATTCCACCAAGTTATTCAGCGTTTAACTATGACT 24358
24376 TTCCCGCCTTTGGTGTATTTGATCACAATATTCGATGTACAGAAAGATACATTTGCGGCT
24436 AACTGATCGCTGTACGTAATCGCGGTGACCAATCTCTGATTTGTTGGTGCGGTTTTG 24442
24496 TT
...
24575 GATAGAAATCTCCATTCGAATCGTAACTGTGAAATGCTTCCGATATATGCGCGCAT 24625
24635 ACTGTTCAATCGGTGATGTAACAGTCCGGCTTTAAGAGAAATTT
...
24815 CCTGCTCAGTTTGAATTATCAGCAGCACTTTTCCGAATACATCAGCAACGCCCTCT 24865
24875 GATGTTGTCGATTAATCTTTTGGGTTCCGGGGGCCAGTTCTGTAAA
...
25107 CGGTGCACTACATTATCTTTTATTTTCTTGCTCTTCCCATTTGTTCAAGCGCTCAA 25157
25167 AATCGGTTTCAGCAGAAACCAATTCAAGATGTTTTTGGCAATCGTTGTCAGACCGTGTGAA 25205
25227 TATCCCTGCGGAACATATAGGTTGCTTTTAAAG
...
25264 GATTACAAATGGGATGCCGTGAGTCTTAACGTGTTGCTGTTTGTATATCGCTGTA 25314
25324 TTTGCGATCTCTTAATCTGCTGAGATAGAGTTAATGCTAC
...
25330
```

```
19258 AGTGATCGTGTGCCACCGCAGCAATATGTTTCCGGCCAGGTAAGTGCAGTCGTAGTT |-----|
19308 GACGGTTTCAGAGAGCGAATAAACCCGGGTGCATATCCAGCACCCTCCGAGATA 20918
19318 CCATATATGTACATCCGTAGTTACTTTCCTACAGATTACTGTATAGACTTATCGCTGCAT |-----|
19368 GATTAAGACCGAAAAAGCCTGCGCACAGGCACAAAAATCTCAGAA 21086
19378 GATTAAGACCGAAAAAGCCTGCGCACAGGCACAAAAATCTCAGAA 21146
... 21206
19579 CATCATCGTGGCAATTGCGGTAGTAGAAGCCGTTTCGCGGATTTGTTATAGCGTTCAAC |-----|
19629 CTGCACCTCATATATCATATGCTTTTGGCGTTGAGGTAAATGGCGAAGATATCCAGCTTTAC 21266
19639 CTGCACCTCATATATCATATGCTTTTGGCGTTGAGGTAAATGGCGAAGATATCCAGCTTTAC 21326
19655 CGGAATCGGCGTCCGATCGCCATCCACAAATTCAGAGCGCGCGGCATCAAAATGGTT 21386
19699 CGGAATCGGCGTCCGATCGCCATCCACAAATTCAGAGCGCGCGGCATCAAAATGGTT |-----|
19706 19714 ATACGTCAAATTTTTCACCGGATCGGATAGTAGACGACGAATTTAAAGACATCATCA 21446
19759 AGCTTCCGCGT 21506
... 21566
19830 CCACCAGCGGTGGGAGATATACAGCCGTTCTTCACTATTTGSAAGATGAATCGTCTCT |-----|
19880 CAACGGGAACACTACTAGTGCAGTGCATTAATTCAGTTCAATGCT 21566
19890 CAACGGGAACACTACTAGTGCAGTGCATTAATTCAGTTCAATGCT 21566
... 21566
20231 TGAACAGCTTGGAGAAATATGAAGAACCGATATTGACAGGCAATTTTAACGTCCACT |-----|
20281 CACACCATTTGGCGTGTGGGGGACGACAAACCCGATCGTCCGACAGCATTCGGGTGATG 21752
20291 CACACCATTTGGCGTGTGGGGGACGACAAACCCGATCGTCCGACAGCATTCGGGTGATG 21812
20315 20328 TCGTAATTCGCACATTCGCGGTTCGCGCAAAAGTTGCGCGCAAA |-----|
20351 AAATATCTACTGACACAGGGGTATCAGTTCATTCGCTTCGCGCAAAAGTTGCGCGCAAA 21872
20411 ACGCTGCTGGGCGACAGAGGGTACGGCACGCTGGCAGACGTTCCGCGAAAAAGTCGATATG 21932
20415 20447 GTCGATGTTTTCCGTAAATTTCAGAAAGCGGCGTGGGGCGTAGCGCAGAGCATATCGCCATT |-----|
20471 GTCGATGTTTTCCGTAAATTTCAGAAAGCGGCGTGGGGCGTAGCGCAGAGCATATCGCCATT 21992
20531 GCGCGAAAAACCTCTGTGATGCAACTGGGGGTGATTAATGAGCAAGCGCAGTACTGSCA 22023
20591 CCGGATGCGCGGGTTGAATGTGGTGAATGATGATGCTCCCGGCTATTTGAGATCCCTGCGCTG 22083
20651 GGCTGGCCAAATAAAAAATCCCC 22143
... 22143
20801 TCGCTACTTAGCTGGGCTTTCAGCCAGGTAGGTATGACACCGGTAGGCGCTTATCGCTCC |-----|
20851 ATCACACAGTGATACACAGGAGCGGCGGAGAGCTGTCATTCACCGCCAGCTCATCAGGC |-----|
20861 ATCACACAGTGATACACAGGAGCGGCGGAGAGCTGTCATTCACCGCCAGCTCATCAGGC |-----|
```

```
16949 AATTTGTACTGCCGGACGAAAAAGTGGCTGTGCACGGTACCGAATGGGGCGAGACGC
|-----|
17009 AGCGTTTAAACATCATCTGTGATGCTCACTGGCGGGGAGTGGAGTGCGAGATGACG
17009
...
17105 TTGCCACGGGCACTGGAGAAAAATTAATGGCTGACCGGTGAGCAAACTCTGGCGTTACAC
|-----|
17155
17165 AACAAATCCGGCAGGCTTTGTCGGCTTGCCTGTCGGTTAACCGACTGAAGAATTGC
|-----|
17187
17225 ATAACTGCCGTGAGGCGGTGGCGTAAATGTCAAGCCCTGTTGAAGATATTGAAGCGCTC
|-----|
17278
17285 GGTTCAGCAATPACCAAGGGGTATACCGAAGCCATGCTTACCGAGTATGC
...
17339 TTTTCCGCGCAGGTGAGTCTTTCACCGCTGAATCCGCGCAGGCCCGGGGAGTCTTAATG
|-----|
17389
17399 GCGAGCAATCTCTGTGTAGTGTGCGAGGTGCAGAAAGCGGAAACGTCGGTGTGTGG
|-----|
17442
17459 CCCGTGCAGGCTGGTTGCTGGCGCGTGGTGAAGCGTCCCTGAGCAAAATTTATTGCTGG
|-----|
17490
17519 CGTTGGTGGCAAAAGCCGCTGAAGATGGAACGACGTAATTCGCAACGGCTACATACCG
|-----|
17522
17579 AAGACATTACCGCAGCAGCAGTTTCATGCGCTGGCGCTGCATATPATTTCAGCAGGCGACGA
|-----|
17610
17639 AAAAAGTCCGATAGTACGACAAACTGG
...
17772 AATGCAGTGCTCAGTGCAGAGTAAGTAACTTCTGGGATGATGAAAAATTACAGCGTCTGT
|-----|
17822
17832 TGCCTCAGCGCCTCATCTGCTGGTAAGTCTGATGGGATGCAAGGTGGTGACAGGACAGA
|-----|
17846
17892 AATGATTGGC
...
17945 AGTTGATGGCCCCGTTATTAAAAAGCTGAAAAGTGCCTGAAGGACAGAAAACGCTGTGC
|-----|
17995
18005 ATTTTGGGGCCTTATTCATCAGCGATTTGATTTCTGAGAGAAAGTGGCTTTATCAGCC
|-----|
18027
18050
18065 CGTGAACCATATTCTGTTGATGAATTTCAGATATCTCGCCGCGAGCGGAGCGTGTCT
|-----|
18117
18125 TAGCGGCATTTACGCAAGCAAAACAGTCAAGACAGCTTGTGCCGTTGGTGTGATGACTGCG
|-----|
18153
18156
18185 AGCGATTTTACCGATTTCAGCGGCTGGCGCAAAATGTCCTCAACCAACGCTTTCATGAAACT
|-----|
18186
18245 TTGTTGAAGGCGCAACGCTGTGATTTAGACACGACTTACCGTTTAAACAGTCTATCGGTG
|-----|
18257
18305 AGGTGGCA
...
18365 TAACCAATGGAACAAAAAAGCCGTCACGCTTATTGGATGAGAGTCACTTGACGCTTTGCG
|-----|
18415
18425 TGGATAGCCTCTCTGTGTTATGCAAAACCGAAGACGCAATCTGTAT
...
18545 ACTTTATGACCATTTGATGCGACAGCAAAAGGCGCAACAGCGGATTACGTCCTCATCGTTGCT
|-----|
18595
18605 TGCAGAGAGGAAGTATGTTTCCGGCTGGCGGCGCGGAGTCAATTATGAAGAGCGGC
|-----|
18660
18665 TACTGCCACCGGTTGAGATTTCCCGAGCGGTGAAGAAAGCGGTTAATGTACGTGGCGC
|-----|
18691
18725 TGACCCGGGGCAGCTCATCGGGTATGGGCACGTGTTTAAACAAAGAAATCCCT
...
18791 TACTGAAAAATCTGGATGTGCCGCTGGCGAGAAAAACGTAAGAAAAAGGTGGCTTGGC
|-----|
18841
18851 ACCTGTGCATATTTACTTTCAGACGGTCCCGCAGATTAACGCTGATATATGGGATCAGANT
|-----|
18886
18911 ATCGACCGCGTCTGTTGAANAATGGGCGACTGGAATATGAAGTCTGCCGTTGCCACGTTGCT
|-----|
18920
18933
18971 GCGCAGCCGGAATGTTCCATPACGCTGCCAGACGACGAAGGCTTTCACGTCAGATGCTG
|-----|
18971
18971 GCGCAGCCGGAATGTTCCATPACGCTGCCAGACGACGAAGGCTTTCACGTCAGATGCTG
|-----|
18993
19031 CGGCACGGCATTTAGTGG
...
19092 AATGGCCAACTGTGTGTCACCCCCCATTTGGGCCACTCAACATGCGGTGACGTTATG
|-----|
19142
19152 CCGGTCCGCGGGAATTAAGTTACCGGTAGTGCCTGTTGCATACA
...
19152
```

```
14227 CAGTTGATGATGTAAACGGGTGATCATACATACCAAACTGCTCATCAAGAAACGGCAGAA 15734
14287 TTTCAGGATTAGTTAACTGCTCCCGGTGAGCACCGAGGGCTGAGAT 15804
... CCCAATAGTGTAAATGATGCGGATCAGCAATGACCGAAGCAATTTGTCATAGCGTTGA 15821
14346 GTATGTTAAGGACACGACCAACCGAAGCGGCTTCGCAATTTGCGGGGTAAACGTTACGCGC 15864
14406 TCGCTGACATATTTGATACCAACGAGCGCACTCAGCATCAGGTTGTGTCATCGCGG 15924
14466 CGGGCAATACGATACGCCAG 15984
... ATGATGAGTTAGCAAAAGCTCCCGCCAGTGGTATCGAGATCGGTACGGCGCGCTGCC 16032
14611 CTGCGCGCCACGCAATGCGACACCAATCAGCGTATCGATTACGGAGGTAAACGTTACTT 16044
14671 CAAAACCTTCACCCAGTAAGTTAAACACAGCTAGCACCAAAAGTGTGATGAACATCGTTG 16104
14731 CATGAGCGTATTGCAAGTTACGGAAGCGCAAAAAGAGCACGCCGTTAAT 16164
... TAGGTTTGACCAAAAGCACTTAGCATAAAGCCATATTAGTATTATTCAAAACGACGA 16165
14790 ACCAGCTGCCCTTCCAGTGTATGGCACAACCAACACAGACAGGAAATGCCAATGGCGTACT 16224
14850 ACCAGCTACCAATATCTTAACTTCAGGGGTGGCGCGGTATAGTTTGGCTGG 16284
14910 CAGCAAAACAATTTGTACAGAGATCCAAATACCGGTATGCAATTCGGTTATCTGATG 16531
14970 ATGGCGTAGCCGAAGACACAACACGACATTTCTACCGCATACGGAAGGCGCGAT 16591
15030 TCCGGCGTGAAGTGACGGCT 16649
... ACGCATAAACCTGCTTTACTGTGTTCAGTATATGGCGGCAATTTGATACGTTT 16709
15333 TGGCAGCGCTGGCCCTGCATCGACATCAGCCGCTGAAAACGGAACAGACAGCATCGCTGG 16769
15393 CGAAATGTTACGCAATGTTTATACTGATATAGAGAGAGCTGG 16829
... ATGGGGGTGATCTTCATATCCGGGATTTAAAGTCTCCGGCATCGCCACGAATACCTTA 16889
15456 TGTGCGCAAAAGTAATATATGACAGCTGCGACAGCGTTCCCGTTGACCACAGATCGCCACGT 16908
15516 AAGCGGGTACGACGAGAGTTTCGTCTGATTTCAATGTGCGCATCAGAGACCGTTGGCG 16971
15576 AGAGCCAAATCGTACAGCGGTGCGTGGCTTTC 16987
... TTCTCTTCAATCAATTAGTGGGATTCACCTGTAAAGCGGCTGTGTTAGGGCGACGCTGC
```

```
11156 GATTGAAGATGCGAAAAACAAAGAAATACGCAATAAAA
...
11290 GCAGCGGGGTTTTTCTACGACGAGAACTTAAAGCTTGGCGTGAATTACACGCTCTTGG
|-----|
11350 ATACCTTTAAGTTCGATCTCTACGACGATCCGGAGCCAGCAGTCGATCAGTCACGA
|-----|
11410 CGCTGTTTCACGTTGTACAG
|-----|
...
11481 GTCTGCCGGGATACCTTTGGAGATCAGTAATCAACAACAGACTGAGCAGCGGCTCGGA
|-----|
11541 CAGACCCCTGGTGTGAAGCCTCAGAACCGATGCGGTGCTTAACCCAGAAACAACTACGGA
|-----|
11601 ACCGCTCTTCG
|-----|
...
11647 AGAGCAGCCTGACCTTCGCGTTTCAGGGTTGCTTTGTGAAGTTGAACAGACGCTCAGAC
|-----|
11707 TTCAGAGTGAAGTCTTGCTCTACTCTCCGTCGCGGACGCTGGAG
|-----|
...
11839 ATGGTGTGTGCGTACCCGATGTTGTGTCCACTGTGATTTCCAGACGGGTAGGGATTTCA
|-----|
11899 GGAAGTATGCGCTACTCAACACCGCCAGCGAGAGACCGGAGAAACGCGGCTGTGTGTTT
|-----|
11959 TTACCATTAACGTTGATTTAGTGTCTGCACGCCATACCATACCAACCCAGACGAGTGTAG
|-----|
12019 ATGTCCAGGTCGTCAGTGAATGGGTAAACCAATTTAGCGGTCAGTTGAACGCCCTGAGCT
|-----|
12079 TTGTATGACCGTTTTTCAACGCTGCTTTGTACGGCATACGACCTAACAGTCGTAAACC
|-----|
12139 ATTTCAAGCCCAACATACGGGTTTAACTGTGTAACCAACAAAGACCAACGCGCCAGTTGG
|-----|
12199 TTTTCATGGGTGCGGCATTTGTTGATGAACACAGTGCAT
|-----|
...
12440 CTGCTCAATATTAACTCTACCGATATCTTGCTTATGCCGACACCCCTGCGCATGTAA
|-----|
12500 AGTCTACAAACGTAAGTTGAAAACTTACAAAGTGAACCTCCGTCAGGCATATGAAAAAAG
|-----|
12560 TCTTGATAAGGTATGTAAATCTTTTTTGTACGCAATATTACAGAAAGAAATCGCGG
|-----|

12620 AAACCGCTTCAGACAGCCTCCGCAAGSAAATTAAGTCACGATGAAGCATTTGGCTTGG
|-----|
12680 CGACAAAAAAGTTCCAGGATTAATCCTAAATTTACTTAATGATACAAATTTAGA
|-----|
...
12756 TGTCGTGGCGTGAGAGATGCGCTTACCGAGCATATATAAACCCCATGCGTTACCT
|-----|
12816 TCATTTGCCCGCATCAACAGTTTCAGCATGCTTTCTTCATGCAAAAT
|-----|
...
13138 CTGTAGTGTTTCCGTAGAGACACCGCAATTTTACTTGTGCGGATGAGACGACGAG
|-----|
13198 AACGATGTGCATAGCCTGAAGTGTACATAATCAATCCAGCCCTGTGAGTTACTGT
|-----|
...
13272 CAGTGACAAACAAGATCAACCCATTTTCGGAAGAGCCTCGCAAAATTTTCTCGTTGGTG
|-----|
13332 ACGGAAACATTAATTAATCTTGCCGCTTAAGATTAAGTTGCCCTA
|-----|
...
13444 ATGAAAAGCCTCTCTTAAGCGGATCTATTAATCACAAGAATACCTGCGAACGTTGGGC
|-----|
13504 ACAATTGAATACGATCATTTGTTGGCAGTTACAGCCTGACCGTGTGACGACGAGGTTT
|-----|
13564 GCGATGTTTCTGATGCTGAGTTGTATCTTCGGGCTTGCTGA
|-----|
...
13708 TGGCAAAATCACTGAAGCTGTGCTGTCTGCAAAATATCTCTCGATGCAAGCCTGAAA
|-----|
13768 GAGAAAAGCAGCGCAATTAACCCGGGAAAGACTGAAGATTTGCCCAATATGTCTTTTCAT
|-----|
13828 CTGGAACCGATTCTCGGGGAGGTGGGATTTAAGATGTACGGGCTTACGTATCTTGGG
|-----|
13888 GCAAAAATGTGTGTTGCTGCACTGCGGACAGAAAAACA
|-----|
...
14027 CTGACTCTCTTAAGCCGAAACAGAGATTTCTCGGAACCTGAGTAATCTGCGCTGAG
|-----|
14087 ACGACCAATCTCAGGCAAGTAATGCAATCAATTAATCAACTTTGTTAGACACAGA
|-----|
14091
...
```

```
...
8655 ACCACACGGTCCATCATACAGCATGTTGGTGTGCGCAATTGGCGGCTTTAGCGCCAAAC |-----|
8715 AGTTCACCGCGACAGAGGCGAAGAGGCTCTTTTGTATAGGATT 8705 8710
...
8845 TCAGTTCCGAATAAACAGTTCAGCCAAACGTAATGGCCATGAGAAAAAGGTGACGCTTC |-----|
8895 8900
8905 CTGTGGCGATGCTTGGCGGATACGTTCCGTGATGTTTGCATCAGCGCTGTTGGCCTTC |-----|
8948 8953
8951 8956
8965 GCCATCCACACACAGATTAAATAAATAGGAGAGTGCATCAGTCACTGATCTACTGCCA |-----|
9010 9015
9025 GATGGTGAATTTGCCTTCTTCTACGGCTTTTCACAGATTCACTGTGAAGACTTAAATGGCG |-----|
9085 AACGTTAGCTGTGGGATAAATGACACTTGTTCCTCCGGTTAACTCAGTCAGTGTGTCGC 9085 9090
...
9173 AAACCCAGGACCGGCTGGGCGGACCGAAGTATGACTGAACCTGTGATAGCGATACTC |-----|
9223 9228
9233 TGATTACACGGGACATCGGGCGAGGCGCTTATCAAGGCGCGAGATTCAGCCATCGAGGCA |-----|
9256 9261
9293 CTATCTCCATCAACTTCACTGTATGACTGTCTCAATATGTCAGGATGTCGAGAGGAGATC |-----|
9332 9337
9353 TGTTCCTCAAGCTGTAGTTCCGACATCAGAACGTTGATGATCATCATCCTTTTCGCA |-----|
9383 9388
9413 TGGATATTGCCGCCCAAGCTCCGCTTTTGGCTTCGATGTGCGTAATTCACCATCGCAATA |-----|
9463 9468
9473 TGCACAAAGCAGCTAATGCGAGAGGTTCCGCCAAAAGCGCGTGAATGACCCGGAATTCAC |-----|
9533 ATGACCGGAAAGGGGTAAATTTGCCCGGATGGCTTCGCTTGGTTTCAATCAGGATTTGC |-----|
9544 9549
9593 TCCTGAA |-----|
...
9638 CGCCATTTCAGGCTGCTGCGCATTAAGTTAGTGTCTCGCGGAGAGAGGTGTCGCATCA |-----|
9688 9693
9698 CACAGGAGGAGGACCTCTTTAACTAGCGGAGATCCAGTGGGGCTAAGCGGAAGTCTT |-----|
9706 9711
9758 TCCTT |-----|
...
9808 ATCCGCTCCCGGGGAGGTAAAGTATTATGTCTGGCGGTAAATGTACCCAGGACACCA |-----|
9858 9863
9868 CTGGTTACTGACTCCGCATTCGACAAATCTGCAGATATCTTCAAAAT |-----|
...
10018 AGGCACAGAGAGGGGAGAGGGCGGCTATCATCAAAACGCAACCCAGTCAAAACGTCGCG |-----|
10068
10078 GTTAAAGATATTTTTCAGCCGCATCCACAGCAGAGGTTGCGCAGCATGTACGCAAAAG |-----|
10081
...
10142 ATGAGAAATACCGCATTTTGTCTGATGACACACAGCAGGCTGCAGGGTAAATGTCCGATTA |-----|
10192
10202 AACTGGCGCAGGACGCCAAAGAGTTGCTCCGCTTCTACCCAGTCGCGCAGCAACTTGC |-----|
10249
10262 GTTAAAGTCGCAAAATTAATCATCTGCACCTCAGTGGGTACGTA |-----|
...
10307 CGGATGAGTGGCGCGAAACCTCATATGATGACCCGCCACCATGTTGGCTGATCGCTTGT |-----|
10357
10367 AGCTAGCGCGGCGCATTTGGCAATTAAGATTCACTGAGACTCTT |-----|
...
10460 AACTGCTCCAGCGCAAAATTGCAAGCCGGCTTGAGTATCATGAAATAAAGATCGTTTTGC |-----|
10510
10520 TCAATCAAAATGTGGCTGAGCAAAATATTCTGATAGTATCGGTAT |-----|
...
10649 TAACCGGCATACAGCACACGTAGAAAGCACCGACAAATACTCTGGCATATGAGGAGT |-----|
10699
10709 CTCACAGGATGAGATTTCTTTCTTCACTGGCTTAAAGCTGATATTCTGTAAGAGTT |-----|
10769 ACACGGTAACATGAGATCGCTATGAAATATCAACAACTGAAATCTTGAAGCGGTTG |-----|
10785
10829 GAAATGGAATGA |-----|
...
10976 GAATCCGGAACGTGGTCAATCGCATGAAAGCAAAACCATTCGGGCAAGAGCAAGCGCATTT |-----|
11026
11036 TAATCGGAACACCGCATACGCGCAAAAATCGATGATCTGAATTAATTTGTGTGCA |-----|
11095
11096 ACGTCTGGCTGTGTCTTGGCAGCGTCCGGTAAGACCTTTCTGAAACGATTTGTACGCT |-----|
11100 11116 11129 11141
```

```
5770 TCGGCGTGGTTGAAGCGCCACACTGGCTGATGATTTGACACAGCTTGAATCAAGGCGC |-----| 7329 7334
      5771 5784 5789
5830 GGGTGAATTC
...
6198 GTGACCAAGCAATGTTCCGCTTCTGGAAGATAGTGGCATTGGGTTTGATCTACGTCAGCG |-----| 7482
6258 GGATGCGGTGGAGATGGGCTCATGTGACAAAGCGTGGTGAAGTGGCGGTGTCAGCTTTGAT |-----| 7542
      6287 6292
6318 GTGCCGGAAGTCTGATTTAGGCGC
...
6502 GTACCGTAAGCAAGTGGCCATTCTTTGGCCGAATATGCTCAGACATTTAAGCATGTT |-----| 7602
6562 ACCGTAATCCGGTCTGATTCTGATTCGAGGACAGAGCGGCTGAAT |-----| 7662
...
6624 GCGGATGTTGTGAGCACCTTGGCGAATTGTTGAACGTGGTTACGCGGATTCGCTGAAA |-----| 7702
      6674 6679
6684 ACCGGAACCTGCTCAGTGGCGCACGTGATGTGATCTCGATTCT |-----| 7762
      6674 6679
...
6770 TGGTTATCAGATTTATCCCGACCGTTAGCGGGGCGCTGGCGCAATTCAGCAAGCATGAT |-----| 7790
      6820 6825
6830 GGAAGCGTTGGATTAAGATCAACAACACTGCATTGAATCCGATGTGAACAGGCAACGAG |-----| 8050
      6833 6838
      |-----|
6890 TACGCTTTCTGAAGTCAAGCGCACATGAAAAACCTGCAACGACGCTGATAGCATGAA |-----| 8110
      6929 6934
      6932 6937
6950 CAAGATCCTCGCCAGCCAGTCGATGACAGCATTTGCCAGCGATATGCAAGTCAACGTTGCG |-----| 8170
      7001 7006
7010 TGAATTGAATCGCAGCATGCAAGGCTTCCAGCCGCTCCGACGCTTACAACAAGATGTG |-----| 8247
      7081 7086
7070 GGGGATATATGCAAGCGCTTGAATCAGGTGTTGCCGAAGACTGCACCGGTGCTGAAGACGCT |-----| 8307
      7124 7129
7130 CAATGAGAAGAGTAAAGCGCTGATTTTGAAGCGAAGGACAAAAAAGATCCAGAGCCGAA |-----| 8367
      7124 7129
7190 GAGGCGCAAAACATGAAAAAGTGGCTAGTACGATTCAGACACTGTGGCTGGCCGATGC |-----| 8427
      7219 7224
7250 AGCTCGCGGAAATTAATAAAACT
...
7279 CCAATTACTGTGGTGCAGAGCGGTACACAAAGTACCAGCCAGCCAGGGAATCGCTGTT |-----| 8487
      8552 8557
      8552 8557
      G 8607
```



```
3206      AGCCGCTCCGTAAGATCGTTTTCGAATAATGGAAGAGCGTTTGTACAGAGTTACGGAAG      |-----|
      3221 3226      3239 3244
3266      CAACTGTGAAAGATTTTTCCTGCCAGGTTCTACGTCGGCAACAATTTGCCCTGATGGG      |-----|
      3302 3307
3326      CCGAATGGCTGGGCAAAACCAACGCTGCTGAA
...
3518      CCAGCCCAAGTTTGGGCTATTTCAGAGACTTCCTGTTCCATCCGAACGGCGCATGACG      |-----|
      3568 3573      3574
3578      -|
      3579      CCAGTACGTGGGCTTTCCTGGGGTGAGCGGAACCGCTTGCTGCGCGCTTGTTCCTC
3638      AAACCAAGCACTTATTGATCTTGACGAGAACCGACACGACATCTTGATGTGGAACGCTG      |-----|
      3674 3679      3691 3696
3698      GAACGTGTGAAGAAGTGAATGACAGACTACCGAGGCAAGGATTTGTGTGTAAGCCACGAT      |-----|
      3756
3758      ---|
      3761      CGTCACTTGTTCGATTAACACCGTTACAGATCTGATCTTGAAGGCGCGGTAAATTT
3818      |-----|
      3819 3824      GGTCGTTATGTCGGCGGTTATCATGATGCCGTGTGACGAAGACGATATGTGGCGTC      |-----|
      3871 3876
3878      AAACAGCCTGCGGTGAAAAAAMACGAAGAACCCGCGCGCAAAAGCAG
...
4087      AGTCTCGCCGATATGCTGCTGCGAGACGAGACGTGAGCAAGCCCTTTGAACCGTGGAT      |-----|
      4137 4142
4147      GTATCTTGAACGTTTAAAAAATGCTGCTGATCGCAAAATATGCTGATTTAAACAGAGACA      |-----|
      4156 4161
4207      GCTGA
...
4316      CATCATCATGCCGGAAGACATCTGTGCTCGAGTGTGACATGCTGTGGCGGTTACCG      |-----|
      4366 4371
4376      CGCCTTGAGCATGCTCAGAAAGCGGCATGTCGCCGGGTGTGGCAACGTTAAACCGTGGCG      |-----|
      4420 4425
4436      TGGGATGCCCGCGGACGCTCCGACCGCCTATGCGTTGGCTGCACGTTCATGCTGTTG      |-----|
      4452 4457
4496      CTGTCCAACCTTGTTCCTTTTGTGAATATGAAACGTTGCAGAGAGTTACAGTGAATTTACA      |-----|
      4526 4531

4556      TTACTGGAATTCGCGCGCTGCTTTT
...
4831      CCCCCTGGTTTATTTTGTGTCCTCAACTGCGCGCCTTTTCAGTCCGTTGATCGTCGCTG      |-----|
      4881 4886      4884 4889
4891      GTTATGGACGATATCGCCCCGATGCGAAGACTGCGCCAGCCGCTTAAACCAAGCGGCTAC      |-----|
      4898 4903      4904 4909
4951      GGGGATACGTACAGGGGCTGCGTTCGTGCTGCTGCTGTTGTACGGCAATCCT
...
5117      AATATTTTGGCCATCATGTGTGACGAGATTATTATAGCTCGAAGATCCATTCGACGATCTC      |-----|
      5167 5172
5177      GCTGGGCTATCTCTGTTATGAGCGAAGGCTTTATCCGTCGCTGCGGTGATCTTCTG      |-----|
      5216 5221
5237      GCCAGTATTATGTGCGCAACGTTAAAGATGATCCGCAATCGCGTGTGTTGGGATGCC      |-----|
      5254 5259      5267 5272
5297      AAAGGCATGCGCAAGCGCGACAGTGA
...
5328      TGCATTTGATTTATGAAGTTGTGAGTTGTAGGCCGCTGTCGATGATGATGATGCTTCG      |-----|
      5378 5383
5388      TTATCGCGGCTCTCTCGCGCTGCTGCTGATGAGGTTTATATGATTTATTCGGCA      |-----|
      5404 5409
5448      TGGGTGATTAATGTTTGTCTTATGTCGATTAATGACAATGTTTCTGTATGACGTTTG      |-----|
      5471 5476      5500 5505
5508      ACCCGGTTTGTCTGTTGGATGTCACCTGAATCAGACATGAGAGTCTGACAGATATG      |-----|
      5526 5531
5568      GAATCTAATAATGG
...
5584      AAGCCAATAATCAGAAAGTGAAGAACTGCTCCCGTGTGATATTTCCATGTCACGG      |-----|
      5634 5639
5644      CGCTCATTTGGGCGCTGGGTTCTTTTATCATTAACAGCATCAGAGGACCGGAAG      |-----|
      5647
...
5710      CCGCAATGCGGAAGAAATTGAAGTGGCAAAACACATTTAAAGCCGTAGGCTGACG      |-----|
      5760 5765      5766
```

```
886 GGGCCGTAAGGCGTGGCTGAGTATAGCTCTCATTTTACGGTTGCGACAGCAGCTC |----| 2057 2062
      936 941
946 GGGTATTAACGTCGACGCACCTAACGCCCGTCTTGGGGGAAVTGGTAACGTATCACCT |----|
      969 974
1006 TTGAGGTGAAGAATGTCGC
...
1040 GCTGCCGAAGAAGCCGTACGGCAGACAGTTAGCAACACCCGCAATACGGGAGAGCTTGA |----|
      1090 1095
1100 CAGCGAACCGCGCTGATTGGCGCTGCATAGCCGTGGGTGCGATCATGAAAAACAGTT |----|
      1109 1114
1160 TGGTG
...
1263 TCGTACAGAAAAATTACCAAGTTGCCGAAGCACCTCCAGACAGCAACCGGGAGTGTA |----|
      1313 1318
1323 GCGGAAGACTACGCCCAACCGCTGCGTAAGAACCTCAAAAAATTGAGAAAGTGGCTCGT |----|
      1313 1318
1383 CAGGAAGGATTGAATGTTACCGCCTGTATGACGGAGTCTGCCGAATATTAACGTTGCC |----|
      1413 1418
1443 GTTGACCGTTATGCCGACTGGGTGGTGGTGTCAGAGATATGCCGCCAAAAACTATTGAT |----|
      1413 1418
1503 GCGCACAAAGCGGTCAGGCTGTGATATATGCTGCAGACCAACATTTGGTACTGGGG |----|
      1519 1524
1563 ATTGCGCAAAACAACTGGTGTGTAACCCGTGAACGTCAGAGGGCAAAAAATCAGTAC |----|
      1597 1602
1623 CAGAAACTGGCGGAGAGGCCAGTTTCTGGAAGTTACCGAATATTAACGCTCACTTGTG |----|
      1668 1673
1683 GTGAACCTGACGGATTATCTCGATACTGGTCTGTCTCGATCAGCGTATCGCCGCGT |----|
      1731 1736
1743 ATGCTGCGCAGATGAGCAAAAGCAAAAGATTCTCAACCTGTTCTTATTAACGGCAGC |----|
      1800
1803 GCCACCGTGCATGCGGGATTAGCGGTGCACGACGACCAACCAACCGTGTATAT |----|
      1805
...
1871 GGAGTGGGAGAACCAACCTGCTGTGAATGGCTTAACCGGGCGTGGCGATGCGCTGAT |----|
      1921 1926
1931 TCAGCCGATGCTGGCGTGGCTGGGTGAGGCAATGAACAGTTC
...
2007 TCTAATCAAAAGCAATGGAAGATGCGTTTGATGTTCAAGCGGATCATCTGGCGTGAAT |----|
      3159 3164
```


gene IEQAREDEVRTLRLHQLLHQN^R
318548. .319390
/gene="ECS1409"
CDS 318548. .319390
/gene="ECS1409"
/note="unknown, similar to hypothetical proteins e.g. L0010
|Escherichia coli| g113414878|gb|AAC31489.1| percent
identity 81 in 111 aa, ydla [Plasmid Colib-P9]
g114512489|dbj|BAA75138.1| percent identity 37 in 265 aa,
L0012 |Escherichia coli| g113414880|gb|A"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34832.1"
/db_xref="GI:13360869"
/translation="MSDITTSRPEVNGHMDVICSTSRHILAVRSKSLIDPLIRO
LAETSVLTESIGGKTAPDMAMKOPRCGCWMEKPEPMKATITRNLDEIRDLMORS
GMUSLMDTQARDWYRSLEVDNPEISEANLSTFEOLHONKDEFERGVINVERGLN
WNKTKNCPKRGSK11VNNLVRMDRWGPHLITGOADRLADLERMLHLFSGKPIPDNR
GNITRLRDDHIQSOGKESYEDEMFSIRYFKKSAHITFRKPELVRDLNDIIAHYPD
MLAV"
complement(319540. .319627)
/gene="serX"
/note="anticodon: GGA, Cove score 68.6"
/product="tRNA-Ser"
complement(319540. .319627)
/gene="serX"
319822. .320799
/gene="ECS1410"
319822. .320799
/gene="ECS1410"
/note="similar to YCDW_ECOLI g111787270 percent identity
98 in 325 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative dehydrogenase"
/protein_id="BAB34833.1"
/db_xref="GI:13360870"
/translation="MRINFAQRFSGVSMIDTFFHTPTDOWMIFALRKKAITQAVRAN
KSGDNDADYALAHWPVEMLAGRDLCRAVFLAGVDSILSKLDHAPMLKPSVPLER
LEDTGMEQMOEYAVSQVLHMFRRFDDYRIOQNSHMOPEYAREDEFTTIGLAGVL
GSKVAOSLQOTMRSLRCMSRTKRSMPGVOSFAGWELSAFLSCQRLVILNPNPEYV
GIINQOLEKLPDGAAYLLNLAGVHVEDDLADLDSGKVGAMLDVFNRRPLPESP
LMOHPRITTPHVAITRPAVAEYISRTIAQLKGRVCCGVNRRANGY"
320854. .321591
/gene="ECS1411"
320854. .321591
/gene="ECS1411"
/note="similar to YCDX_ECOLI g111787271 percent identity
99 in 245 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34834.1"
/db_xref="GI:13360871"
/translation="MYVDLHMHTVASTHAYSTLSYIAQAKOKGKIKLFAITDHPDM
EDAPHWHFTNMRTPRVVDVGILRGIENIKNVDEIDCSGKMFSLDILIAGFHE
VPAPHDKANTQAMITATISGNVHII SHGNPPEYEDVKAVAEAAKHQVLEINNS
SFLHSRGSDEDCRAVAADAGGWAVALSDSHTAFTMGFEBCILDLADVPPER
ILNVSFRLNLFESRGMAPIAEFADL"
321615. .322169
/gene="ECS1412"
321615. .322169
/gene="ECS1412"
/note="similar to YCDY_ECOLI g111787272 percent identity
99 in 184 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative oxidoreductase component"

gene /protein_id="BAB34835.1"
/db_xref="GI:13360872"
/translation="MNERSIICRVLGSLYRPOPDLLVPLFTLLIREGKLAAMPLEQ
DELFLRLQSCDMOVADYNALFDEBCAVPYSAWVEDATEAVRAFLSERGMPL
ADTPADHIGTLILASWLEDSOTFEDESSEALFTLFESEYLLPMCGALGKVEAHATPFW
RTMAPLITDAISAMDELEEDSEE"
322223. .322762
/gene="ECS1413"
322223. .322762
/gene="ECS1413"
/note="similar to YCDZ_ECOLI g111787273 percent identity
98 in 179 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34836.1"
/db_xref="GI:13360873"
/translation="WCSISFMAFSAIRMGNNILSTAITTGILSGIMGWAVSLGL
SWAGLGCTAYFACPOGGLKLAISAATLISGVNAWVITVGSALAPHLIELSYVTG
IVAFIMCIOARQOLLSFVPGTFIGACATFAGCGDWKLVLPALGLITFGAMKNSGLW
LAARSAKTALRHOETIKKA"
complement(322826. .323659)
/gene="ECS1414"
complement(322826. .323659)
/gene="ECS1414"
/note="similar to CSGG_ECOLI g111787274 percent identity
100 in 277 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curl production assembly/transport component
CsgG"
/protein_id="BAB34837.1"
/db_xref="GI:13360874"
/translation="WQRLPLVAWVLLSGCLTAPPEKRAPPLMRAOSYKRLTHLPA
PTGKLFVSVYNIQDETQGFKEYPASNSTAPQSAATMLVTLKDSRWFIPLERGLO
NLNLERKIIIRAQENGTAIVNNRIPLOSILTANINVEGSIIGESNVSGVGARYFG
IGADTQYOLDIAYNLRVNVSTGEILSSVNTSKTILSYEVOAGVRFIDYORLLEGE
VGYTSNEPVMCLMSALETGVIFLINDSIDRSLMDLQKAEQNQNDILVYRMSVPE
S"
complement(323686. .324102)
/gene="ECS1415"
complement(323686. .324102)
/gene="ECS1415"
/note="similar to CSGF_ECOLI g111787275 percent identity
100 in 138 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curl production assembly/transport component
CsgF"
/protein_id="BAB34838.1"
/db_xref="GI:13360875"
/translation="MRVNAHVALLMLISPLSWAGTMTQFRNPNGMPPNNGAFLLNS
AQAOQSYKDPYNDPFGIETPSALDNFOALQOSQILGGLSNIINIGKPGRMVINDYIV
DIANDRGOLQNVTRKTGOTSTIOVSGLQNNSTDF"
complement(324127. .324516)
/gene="ECS1416"
complement(324127. .324516)
/note="similar to CSGE_ECOLI g111787276 percent identity
100 in 129 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curl production assembly/transport component
CsgE"
/protein_id="BAB34839.1"
/db_xref="GI:13360876"
/translation="WKRLRMIVAAEPLPAAGNLHAEVEVPGILTDHTTSSIGHDFY
RAFSDKWSYDTGNLTINERSARGSWITTTVNQDVIFQTFLEPLKAKDFEKTVAFL
IQTEALNRROIINALSTGDLAHDEF"

```
gene
CDS
    /product="hypothetical protein"
    /protein_id="BAB34823.1"
    /db_xref="GI:13360860"
    /translation="MGEMNEVYCLRSPSLANHVLSCLIPASPVLQTFESIADNPETICSLVVFSSPKIK"
    /gene="ECS1401"
    /note="unknown, similar to hypothetical proteins e.g. YfjQ [Escherichia coli] g11723629|sp|P52132|YfjQ_ECOLI percent identity 73 in 271 aa, YafZ [Escherichia coli] g11249548|sp|P72061|YafZ_ECOLI percent identity 73 in 271 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34824.1"
    /db_xref="GI:13360861"
    /translation="WRLASRFGYANQIRRDREPLTHEELMRHVPSTFGENRHTSRSEHYAVIPITVLENIQDEQFQPFACQTRVDOSRREYTKHMLRLRAGQITGQHVPEILNSHDSSSYOMLPQYFRAICTNGLCVCSGELPVRGNVADVIEGAYEVGVFDRIEERKDMOSLVLPAPRALQALYRYGDEHPVTADILTPRRRDYGRDMSAYOTIOENMKGISGRSAKGRIRHTRAIHSDTDIKLRALWMAETLLESMP"
    /gene="ECS1402"
    /note="unknown, similar to hypothetical proteins e.g. YafK [Escherichia coli] g11249548|sp|P56761|YafK_ECOLI percent identity 71 in 144 aa, YfjX [Escherichia coli] g11723636|sp|P52139|YfjX_ECOLI percent identity 75 in 137 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34825.1"
    /db_xref="GI:13360862"
    /translation="MTIVSHNSTTPPSVSTTASGNPNPOLVATVPVDEQRISFMPQHFGLIPQWVLEPRVFGMDRLCEDYCGGIMNLTYLNGAFAPAPEDDDDEFWVLFNAMNGRNAKSPERAGIAACLMYTSIHACRTGECYAMVHTYRLRDYALQHPESALMRTID"
    /gene="ECS1403"
    /note="probable DNA repair protein (RadC family), similar to putative RadC family proteins e.g. YKfG [Escherichia coli] g113025218|sp|O47685|YKfG_ECOLI percent identity 81 in 158 aa, YeeS [Escherichia coli] g113025155|sp|P76362|YeeS_ECOLI percent identity 98"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative DNA repair protein"
    /protein_id="BAB34826.1"
    /db_xref="GI:13360863"
    /translation="WOOLSPFLGEMTPERSLITRALQTLDRHLHEPGVAFSTRRAAREMLINMAGLREFRVLVYLNNOQLIAGETLFGTINRTVEVHREYIKRLLYNMAAVVLAHNHSGEVTSKADRLITERLVQALGLVDIRVPHLITVGNOVFSFAHGILL"
    /gene="ECS1404"
    /note="unknown, similar to hypothetical protein YeeT [Escherichia coli] g113025156|sp|P76363|YeeT_ECOLI percent identity 97 in 73 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34827.1"
    /db_xref="GI:13360864"
    /translation="MRITRGEMRIHQHPASRLFPCTGKRYRWGSAEATYGREVO DTPGVLAFAERKDSFGFYRVMSVTLN"
    /gene="ECS1405"
    /note="unknown, similar to hypothetical proteins e.g. YeeU [Escherichia coli] g113025157|sp|P76364|YeeU_ECOLI percent identity 89 in 118 aa, YfjZ [Escherichia coli] g11723638|sp|P52141|YfjZ_ECOLI percent identity 66 in 98 aa, GrG start"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34828.1"
    /db_xref="GI:13360865"
    /translation="MSDKLSGITHPDDNHDRPMWGLPCTVRPCGARLVQGNRLHYLADRAGIRGRFSDADAYHLDQAFPLMKQLEMLTSGELNPRYQHTVTLNAGLGCCEADTLGSGYVYLAVYPPAPATTS"
    /gene="ECS1406"
    /note="unknown, similar to hypothetical protein e.g. L0007 [Escherichia coli] g113414875|gb|AAC31486.1| percent identity 93 in 124 aa, YeeV [Escherichia coli] g113025158|sp|P76365|YeeV_ECOLI percent identity 87 in 124 aa, Ykfi [Escherichia coli] g113025213|s"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34829.1"
    /db_xref="GI:13360866"
    /translation="MKTLSDTHVREVSRCPSVTIMQTLIRLLDQHYGLTNDTPFADERYIEQHIEAGISLCDAVNLFVEKYALVRFDDQFSTCPRSQLNSIDLRRARRATGMTNDNYRTVNNITLGRKPEAK"
    /gene="ECS1407"
    /note="unknown, similar to hypothetical proteins e.g. L0008 [Escherichia coli] g113414876|gb|AAC31487.1| percent identity 94 in 163 aa, YeeW [Escherichia coli] g113025160|sp|P76366|YeeW_ECOLI percent identity 65 in 55 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34830.1"
    /db_xref="GI:13360867"
    /translation="WMKLALTLADSVVQALNMGRIVVDVGIELAEILNMWCDNGYSLRVVDESDQTSACCTPPFATVLTGICSTAITTEIDNMLYSLSHQSDFESEMIHFTNGCYLRTDQWSPVLRKLKRLGSKTFRRLVVTLLIRYGVSLIHLDAASACLGCLPFDW"
    /gene="ECS1408"
    /note="unknown, similar to hypothetical proteins e.g. L0009 [Escherichia coli] g113414877|gb|AAC31488.1| percent identity 87 in 65 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34831.1"
    /db_xref="GI:13360868"
    /translation="WKSILTETALDILLIAMLQNDINIDCGSGIIFDNDKDTDSALLPC
```

```

repeat_region      305027..305075
                    /note="IS2-1, truncated (3')"
```

gene

```

                    complement(305230..305412)
                    /gene="ECS1394"
```

CDS

```

                    complement(305230..305412)
                    /gene="ECS1394"
                    /note="unknown"
                    /codon_start=1
                    /transl_table=11
                    /evidence=not_experimental
                    /product="hypothetical protein"
                    /protein_id="BAB34817.1"
                    /db_xref="GI:13360854"
```

gene

```

                    /translation="MYNCGSESYEDYISLSDENYDDDYRERPCGPEVGYDNTEDN
DEASDYALKINEMSD"
                    305741..306613
                    /gene="ECS1395"
                    /gene="ECS1395"
                    305741..306613
                    /gene="ECS1395"
```

CDS

```

                    /note="unknown, similar to hypothetical proteins e.g. yfjp
protein [Escherichia coli] g11744953|pir1|B65042 percent
identity 49 in 289 aa, yefp protein [Escherichia coli]
g112495624|sp1|P63591|YEFP_ECOLI percent identity 95 in 183
aa"
                    /codon_start=1
                    /transl_table=11
                    /evidence=not_experimental
                    /product="hypothetical protein"
                    /protein_id="BAB34818.1"
                    /db_xref="GI:13360855"
```

gene

```

                    /translation="MNPSDAIEALEKPLSSLPVLSRHILEHLRLKLSHEPVIGIMCK
SGAGSSSLCNALFQGEVTPVSDVHAGTRREVRERFLSGHSMVITDLPVGESEDRPA
EYEALYADILPELDLVLMILKADRALSYDEYEMRHILRHGHOOVLFPVVOADKTEPC
HEMMDAGIOPSPQEOINIREKTDAVFLRPVHPVAVASARTGEIDLVASLMTAPR
DHASPLMTRLQDELCTESYWGQAREQFTGAVDRIFPTVESVCAVARYVLRVROT
VSVARAVMWIIF"
                    306985..309834
                    /gene="ECS1396"
                    306985..309834
                    /gene="ECS1396"
```

CDS

```

                    /note="also similar to Adhesin AIDA-I precursor
[Escherichia coli plasmid pIB6]
g115437881|sp1|Q03155|AIDA_ECOLI percent identity 23 in 678
aa
also similar to glycoprotein [Escherichia coli strain
H10407] g115305639|pb1|AAD41751.1| percent identity 34 in
608 aa,
probable adhesin,
similar to outer membrane fluffing protein [Escherichia
coli] g117466262|pir1|G64964 percent identity 68 in 927
aa"
                    /codon_start=1
                    /transl_table=11
                    /evidence=not_experimental
                    /product="Aida-I adhesin-like protein"
                    /protein_id="BAB34819.1"
                    /db_xref="GI:13360856"
```

gene

```

                    /translation="MKRHNTSYRLVYNHITGLVASELARSCKRAGAVALSIA
VTSVPALAAKVVQAGETVNDGTLTNHNDQIVFETANGMTISTGLGPOSEEWTGQO
WIONGGJAGTMTVTYNGROYVLREGTASDVYIRGGGOSLNGLAVENTTLNNKGOWH
EGGVANGTITINDGYOVSKSGGLATGTIINTGAEGSPDSNGYTGQVQGSTAETTTN
KNGROIILFSGLRDTLIVAGDOSYHRAVLANPTLNGYOVYHNDGLAINTVINEGGM
OYKAGGAAGNTTINONGELRYHAGGEATVYQNTGALVYSTRATYIGTRNRCNPLY
ENGRKADVLESGRDLVLESHSNQNTLVDDGGLAVSAGKATSVITISGGALIAS
GATVEGTNAGSKRSIDETSGQASGLILENGSFTVNAAGQAGNTVGHRTLLAAGS
SLSGRTDLSKSGAMVINGDVSTGDIYNAGEISFTNQTTPNAALRAVANSNPFVTH
KLTFTTNLTGOGGTINMRVRLDGSNASPOLYINGQAGTKTMLAFTNNGVSNLGVATFG
OGIRVYDAONGATTEGAGFASRLPLOGAFNNTLINRSDSDVYLRSENFARAVPLTYT
SMLTQAMDYRITLAGRSRSHOTGVNGENNSVRLISIQGHLCHDNNNGTARATPESSSS
YGVRLLEGDLIRTEVAGMSLTTGVYGAAGHSVDVKKDDGSRAGYVRDDAGSLGGYLN
LVHTSSGLMADLIVAGGTRHSMKASDNNDDRARGMWGLSLETGLPFSITDNLMLPEQ
LOQTWGLISLDGDDGNAGYVFKFGHSGAOHRAFRGLSGSHNDMTFEGECTSSRDTLRDA
KHSVSELPMVMWVQPSVIPTFSSRGDMSMGTAAAGSNMTPSPSRNGTSLDQLGLEAR
```



```

/gene="ECS1377"
/note="unknown, weakly similar to sensor-regulator protein
HutT [Rhodobacter capsulatus] g11075537|pir||A49938
percent identity 33 in 97 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34800.1"
/db_xref="GI:13360837"
/translation="MTRRIDYLDKYCTCFERNESPLRRQMDVLEECROTEGAPPEER
LRLLALNVYVSFELPFRLLRTLPOLJAALREWGISOKNVFNDKRCVYSLSKA
SLSGVDTRYRHLSHRIRRYGMENTSLPQYARREKAPRERIKYALAEGLVLTAD
GLRFGSORIAADVRLRRKAGMVFVTYEVHDLNGTRKKVPAVHL"
293704..293937
/gene="ECS1378"
293704..293937
/gene="ECS1378"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34801.1"
/db_xref="GI:13360838"
/translation="MARSDYDIINLSLEHLENEWLAERGYAGLVDRNRRLAEVYTRKL
ODSFYINVRDALNTAYSEHPEWFSGLVSGDEN"
293948..294163
/gene="ECS1379"
293948..294163
/gene="ECS1379"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34802.1"
/db_xref="GI:13360839"
/translation="MLLSGSGVINGRDFVLECFPPGFGFESQKVVHRSHAGMRLAVLH
KHHRGPIINQGGCKRTMIVFVSLIDFR"
complement(294773..296042)
/note="15629-03"
complement(294775..295662)
/gene="ECS1380"
complement(294775..295662)
/gene="ECS1380"
/note="transposase, similar to transposases e.g. [insertion
sequenceIS629] g117443863|pir||T00315 percent identity 98
in 295 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase OrfB protein of insertion
sequence IS629"
/protein_id="BAB34803.1"
/db_xref="GI:13360840"
/translation="MPLDLKLEHGVGVPCSELHIAPISTYYHCOOORHHPOKRSARA
QRDDMLKREIQRYVDENHOYGVYKRWROLREGISVARTVARLMAVGLAVLGG
KVRTTVSRKAVSAGDVRNRFVAERPDQWLADFTVSTWQGVYAFITIDVAGCIV
GMRVSSMTEFTVLDALQOLMARPSGVVHSDKSOYISLAYTERLKAUKLAASTG
STGSDSYDNAMAESINGLYKAEVYHRSKWRARAEVLELTWVWYNNRRLGLGHIP
PAAEKAYVASINDDLAA"
complement(295662..295988)
/gene="ECS1381"
complement(295662..295988)
/gene="ECS1381"
/note="transposase, similar to hypothetical proteins e.g.
[Escherichia coli plasmid pO157 insertion sequence IS629]
g117444868|pir||T00241 percent identity 96 in 108 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase OrfA protein of insertion

```

```

sequence IS629"
/protein_id="BAB34804.1"
/db_xref="GI:13360841"
/translation="MTKMRPESPEVRORAVRWYLESGEYDSOWAALICSIAPKIGCP
ETLRVWRQHERDPTGSGDGLTTAERQRLKELERENRELRSDIILRQASATRAKAE
DRLMKR"
296481..297215
/gene="ECS1382"
296481..297215
/gene="ECS1382"
/note="HecB-like protein,
N-terminal half is similar to N-terminal parts of
hemolysin activator protein HecB [neisseria meningitidis
MC58] g117227016|gb|AA642103.1| percent identity 34 in 181
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HecB-like protein"
/protein_id="BAB34805.1"
/db_xref="GI:13360842"
/translation="MSTLQNRVADHGVTYRVLRVLSODLKSGLRLVITPGVRYRL
TPSDDYIQLYSSFPARHESLDRDIEGSLDIGNSRIGQRTLELNATSGNLSTQNRQ
LSADTLSTARAGQFSSNGGTINADTLQISAQSLSNRKSGLIQGTGDFSLFSGVDN
REGILLANGAVRLDALSLDNRKRGVOAESPSIQKSPPTFLKPFVAGVCAALLAVSA
IPGMQFLTOPSPPEOHFTWNGCKKO"
297456..297971
/gene="ECS1383"
297456..297971
/gene="ECS1383"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34806.1"
/db_xref="GI:13360843"
/translation="WVTICDGENKTHKTIPTDWTMAGNDREVMLTCHFRSGRRKYTRP
LSVCOITPYVNLNRVLEKRGNAVTSRAELVITYGKVAAYYREEREPYIMKTGID
FOCCSAFTEHAYVNYLCRVANERITFARGNRNHIDENILQIKIYVHPPTALHAACS
GOSKNYIRHGV"
298199..300136
/gene="ECS1384"
298199..300136
/gene="ECS1384"
/note="unknown, weakly similar to part of tRNA-splicing
endonuclease positive effector [fission yeast]
g117493527|pir||T740065 percent identity 22 in 531 aa, also
weakly similar to hypothetical proteins e.g. [Aquilifex
aeolicus] g117514764|pir||D70476 percent id"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34807.1"
/db_xref="GI:13360844"
/translation="MARLSTENRROOFESNISRSSEVLDPDSANAIDVYLOOVKX
LMAINOVASIKARINLENIYKYLQOMOSONLRPEELFSKRYFSSOKTDDLMAIYH
LSDRRIGFRNRIDLLNPLMLKAPLIPERRIALFTSIDLSTYERKIRRQKJSLNEX
EEAFKNSDFKILGLRTWSMILKOLLRNVSTRSSFSSEYRDEDFRITKPPYITG
SSTHSIINSIGKALLDYVIDEASODIYPGILIGCAGANVYVGRKQLPHPVYL
PNSPSPAEYVNCSEKYSILDSVCMLEFRNMPVLLKEHYRCHRKIIQCKRQFYDNL
IPLTVSGEASISLVITAKGNHTRNFSNLELSLEGHYDESSRGYIAPYNAOYVL
AEKVLADPAPYKSVHKRQGRCEDEIYEVLDKRSQHSRNTAFVNPPLVAVAVSR
ARNKFTLVYTGNDVFERHAGHIALIRIKYIYADDEGFESPEVYISAPDILSEYDKSE
RLNSRLNSNDSHFRKSDQIVACLLDLISDSYSMMFHSDIATNQLLLEGGFTTHE
QLTRIDNIEGKLAFPSGLTG"
300180..300320
/gene="ECS1385"
300180..300320
/gene="ECS1385"
/note="unknown, TTG start"

```



```
/db_xref="GI:13360827"  
/translation="MCHRAHOSCTRDVALDNRHLPDCEIDLEKVIWASDFVRNRCV  
YDPEARYPAGNWPYTGOF"  
/gene="ECS1368"  
complement(288048..288194)  
/gene="ECS1368"  
complement(288048..288194)  
/note="unknown"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34791.1"  
/db_xref="GI:13360828"  
/translation="MAFTDQLVKFPVGHGTGDIIVHDQGIIGTKYAVVACCSQHG  
QNVL"  
complement(288170..288478)  
/gene="ECS1369"  
complement(288170..288478)  
/gene="ECS1369"  
/note="unknown"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34792.1"  
/db_xref="GI:13360829"  
/translation="WTNOSGRDNPNLNGRRVQOEDEREYRHOQOGRRVQYORWHFRAH  
CORTGITTRFMILPRIPRFVATSTALLHFHCSTSLYQRLRLPMLGKWPSTLNL  
"  
288993..290117  
/gene="ECS1370"  
288993..290117  
/gene="ECS1370"  
/note="possible glucosyl-transferase, similar to  
glucosyl-transferases e.g. [Salmonella typhi]  
g1|7467230|pir|t30292 percent identity 72 in 366 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="putative glucosyl-transferase"  
/protein_id="BAB34793.1"  
/db_xref="GI:13360830"  
/translation="MKRIIFLTPPLLYGLPLISLAQAFVYIGHDVYISSAKGFANK  
AAEAGLVFDVPGLDSEAGIRHOELERKSNITGFESFSDMADNLIDRAGKRRPD  
LIVPPLDPAGPLAAKRIPIRYMLAVGFAHTSAHIQMLNSNAYRRHVGSGFICD  
LAWIDVAPPSILKNAGEPIYSNRYIPYNGAVETWMDSDRRLISLIGYKPKM  
VDGELIISMVDSANEVDADIILQLAINARFGIRKLPSNVRLLVDMIPKGVFLNADGCF  
IHGGAGNTLTAIYSGIPIQYVFGGADRSVNAEIVAKKCCGIIPDKHGLTSDLVNRL  
YDSTRFCSDOYAAEMAEQPSPAETAEVLMKRLKNGKQL"  
complement(290259..290534)  
/gene="ECS1371"  
complement(290259..290534)  
/note="TTC start  
probable ferric enterochelin esterase (partial),  
similar to C-terminal parts of ferric enterochelin  
esterases e.g. [Salmonella enterica] g1|  
2738250|gb|AAC46181.1| percent identity 66 in 68 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="ferric enterochelin esterase"  
/protein_id="BAB34794.1"  
/db_xref="GI:13360831"  
/translation="MESLPENNTSMVSPQILSAPPKDVNIRLGVSLGCTTVSHVQRL  
HSLIAGLESNLTVYTGCHDYAMWRGAILDELANYNCRKYSNNNSV"  
complement(290521..291288)  
/note="ISI  
complement(290536..291039)  
/gene="ECS1372"  
complement(290536..291039)  
/gene="ECS1372"  
CDS
```

```
/note="transposase, similar to transposases e.g. [Shigella  
boydii] g1|2197010|gb|AAB61273.1| percent identity 100 in  
167 aa"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="transposase"  
/protein_id="BAB34795.1"  
/db_xref="GI:13360832"  
/translation="MSROCTHYGRNPQHGFTSLKKLRPOSVNRIOPGSDVIVCAEMD  
EQMGVYGAKSRRQRLFYADRIIRTVNAVPERPLATLERLSLSAFEVYVWMTDG  
WPLYSRLKGLKLVHSKRTQRIERHNLRLQHLARLGRKSLSFSSVELHDKVIGHY  
LNKHVQ"  
complement(290958..291233)  
/gene="ECS1373"  
complement(290958..291233)  
/gene="ECS1373"  
/note="possible repressor, similar to Insa proteins e.g.  
[insertion sequence IS1F] g1|124915|sp|P19767|ISA2_ECOLI  
percent identity 98 in 91 aa, GTG start"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="Insa protein of insertion sequence IS1"  
/protein_id="BAB34796.1"  
/db_xref="GI:13360833"  
/translation="MASISIRCPSCSATEGVYRNCKSTAGHORVLCSPCRKTWQLQFT  
YTA5OPGKHOKITIDMANMGVCGRASARIMGVLANVLRHLKNGSR"  
complement(291471..291929)  
/gene="ECS1374"  
complement(291471..291929)  
/gene="ECS1374"  
/note="unknown"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="putative membrane protein"  
/protein_id="BAB34797.1"  
/db_xref="GI:13360834"  
/translation="MKIYTRKRRIPELALPSPSLRKKMFTDANVYASSHIEYMGACV  
SPISMWLISRYVDLPLPNODFPFGDIISLCYVCSLIIMQFSYVYVAKKRIHLIVR  
CKETFTISERLIREADREYELMRPVRKRFSTIVFLVLLGCIHSLITTIK"  
291994..292185  
/gene="ECS1375"  
291994..292185  
/gene="ECS1375"  
/note="unknown"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34798.1"  
/db_xref="GI:13360835"  
/translation="WTSPTFLIEFDNDNRDNNHYQLIVSVILTYITLISFNQVIDLKMR  
VCLMWYCVYVLPGLICSYIN"  
292387..292896  
/gene="ECS1376"  
292387..292896  
/gene="ECS1376"  
/note="unknown, GTG start"  
/codon_start=1  
/transl_table=11  
/evidence=not_experimental  
/product="hypothetical protein"  
/protein_id="BAB34799.1"  
/db_xref="GI:13360836"  
/translation="MQITELIISPEGLIRRFVQQAVDHMPNLLAFHFLYSABEIGYG  
QIQFCSFHRVHERITENHNHTVSPAPVLRWLREQHEGAQIRCLLLSOTSICHP  
RGVMADECAQVLDLLQOTWSVISAQGCQVRCFRVARRGSSGGQVALKTAQNSM  
SOVATITIR"  
292985..293608  
/gene="ECS1377"  
292985..293608  
CDS
```

```

gene      /translation="MNLOSQNIPLQOSTRLNLQYPAKSGFKGEPDTCLEMLNAQK
VSDSDSDFIFYNLNLSSPEGAVRLVYGSQASIEIALDRVPANVSKIAITVVIDGDDTIS
"
CDS       complement(280747..280971)
          /gene="ECS1359"
          complement(280747..280971)
          /gene="ECS1359"
          /note="unknown,GTG start"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34782.1"
          /db_xref="GI:13360819"
gene      /translation="MFGQNNSPMLRAGOVLTILITWRSHVSGSIPLSFATDQAVHH
CRTIARHLKYLRSQOEGIIISRHSRGPRV"
CDS       complement(281205..283295)
          /gene="ECS1360"
          complement(281205..283295)
          /gene="ECS1360"
          /note="adhesin, similar to Iha adhesin [Escherichia coli
          O157:H7 strain 86-24] gi|7108480|gb|AAFP6432.1|AF126104.1
          IHA_ECOLI percent identity 99 in 696 aa, also similar to
          exogenous ferric siderophore receptor R4 [Escherichia coli
          strain CFP073] gi|3661500|gb"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="Iha adhesin"
          /protein_id="BAB34783.1"
          /db_xref="GI:13360820"
gene      /translation="MRITPLASVVICPGFSASIAAEDVMIVSASGEKKLTNAA
SVSIVSOELOSQYHDIAELRSGEVDSGKGGLGISIRGMPASVTLIDG
VRGSSDYTPNGFSAMTGFMPPLAIEIEVIRGPMSTLIGSDAMGAVNITTRN
ADKMLSSVNAGLNDQSNKMGNSQEFNMSGPLVDVSLQYRGSTQOQSSVLS
SDTAARTIRPYTESQNTLGAFLDKWASQEDVLMFDMDTTORYDNRDGQLSLTGYS
DRLTARENKISAGYDHTFETFTGWSYLMNNETENKRELVRSYLKDKMGLAGOPRE
LKESNLNLALTPLGESHLVTVGGFQSSSMKDGVLASTGETFPKQSMYSVAEDE
WHLTDLALTAGSRYEHHQFGHFSPRAYLVMDVADAMTLKGGVTTGYKAPRMGOIH
KGISVSGCGKTNLGNPDKPRESVYAGVYDNRAGLANAVTGMPTPFSKIYYS
SINDNTNSYNSGKARLHGVFACTLPWSEDDVTLNITWTNSEQRDGNKGAPLSY
TPEHVNAKLNWQITEVASWLGARGRKPRFTQNTSSLSAVOKRYDDEKGEYLKW
TIVVDAGLSMKMTDLTLNAAVNMLLNKDYSDVLSAGKSTLYAGDYFQNGSPTGYV
IPERNYMSLNYQF"
gene      complement(284099..284221)
          /gene="ECS1361"
          complement(284099..284221)
          /gene="ECS1361"
          /note="unknown, similar to part of hypothetical protein
          [Shigella flexneri] gi|5880472|gb|AAD54665.1|AF097520_3
          percent identity 44 in 40 aa"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34784.1"
          /db_xref="GI:13360821"
misc_feature /translation="MSGKTFEFDVICFVYFITIVCKSFKLIPKAKGFRK"
284269..284565
          /note="partial CDS, similar to part (117-213 in 347 aa) of
          Shia [Shigella flexneri]
          gi|5532447|gb|AAD44731.1|AF141323_2 percent identity 72 in
          97 aa"
gene      284694..284966
          /gene="ECS1362"
          /gene="ECS1362"
          /note="ECS1362"
          /note="ECS1362"
          /note="unknown, similar to C-terminal part of Shia
          [Shigella flexneri] gi|5532447|gb|AAD44731.1|AF141323_2
          percent identity 49 in 73 aa,TTG start"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental

```

```

repeat_region
gene      /product="hypothetical protein"
          /protein_id="BAB34785.1"
          /db_xref="GI:13360822"
          /translation="MPGMMYVNSVLSKERSEMPKYMMLMAEDDKVCELIFKQDN
          LILKINRFPIITIKKTYTLOGKYYWMLIFVLEMLIVCTLGANS"
CDS       complement(285988..286363)
          /note="IS1151, truncated (5' and 3')
          complement(286087..286239)
          /gene="ECS1363"
          complement(286087..286239)
          /gene="ECS1363"
          /note="unknown"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34786.1"
          /db_xref="GI:13360823"
          /translation="MLYDAFTVPCQRPRIWIRVGNIELQALGVKAPMSSTRYAA
          ORKRSF"
misc_feature complement(286244..286408)
          /note="partial CDS, similar to part of IS150 putative
          transposase gi|790866|gb|AA96489.1| percent identity 40
          in 52 aa"
          complement(286715..286909)
          /gene="ECS1364"
          complement(286715..286909)
          /gene="ECS1364"
          /note="unknown"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34787.1"
          /db_xref="GI:13360824"
          /translation="MNEENKIKRLQLQARQREFATRWGRAALPHERLTYPLEDLK
          NNGKKIRSDKKHFELECEDK"
gene      286989..287213
          /gene="ECS1365"
          286989..287213
          /gene="ECS1365"
          /note="unknown, TTG start"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34788.1"
          /db_xref="GI:13360825"
          /translation="MELSTKGEARLYRNGSMALMPVCCSIMASLSIRCCSSRKMPA
          SSGDKSVITRKAMRCGALRADGSVPFY"
CDS       287229..287501
          /gene="ECS1366"
          287229..287501
          /gene="ECS1366"
          /note="unknown, GTG start"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34789.1"
          /db_xref="GI:13360826"
          /translation="MSYOPGLGVLPDLSSRGYLSKNASSNRFGLVNGDGLSKDSRCM
          CTFSSRMSSMDNAIACSSSLVICTWAIYRSHYCKNVLLQLA"
gene      287806..288000
          /gene="ECS1367"
          287806..288000
          /gene="ECS1367"
          /note="ECS1367"
          /note="unknown, TTG start"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34790.1"

```

```
gene
CDS
    /protein_id="BAB34773.1"
    /db_xref="GI:13360810"
    /translation="MNKKIFMEGLSSORDIIGCVKSPQAQNNFATVFASHRNERNE
    ILVADISLLEDEPOKRLQFIQETIQYGIHIIHTGNSQWFEHRSAIPTGATLT
    TGAIVGVMLTLADEKVTFAQFMQKGLPVPSMWNTLAEKTHLAAPFTDSVCVK
    PVTGIYGVGFWRPDDASPMAYFNHPEHRLVSPQOYTAASAASFKPLVMPYLPGP
    EFSVIDLADKEIILAAVGRREGAIOYLVNMGSAAMELACDCAKMAKADGIVNCTRD
    VSGNVLLETMRPSSGCVYTLHSGVNLPGLEPAFKGLMSDMDVROSAKNTFSPAV
    RSTIDVYAIPESSNLNLN"
    /note="possible tellurium resistance, similar to Ter2
    proteins e.g. [Serratia marcescens] gi|6094454|sp|Q52353|
    percent identity 98 in 193 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance protein Ter2"
    /protein_id="BAB34774.1"
    /db_xref="GI:13360811"
    /translation="MWSLVKNQTVSLSESSALSQHLHGLGWDPVKRRKGLGLFGGN
    DSIDLDAQCVLMDSTGKTIDTIFMRKLKSCGAVVHSGDNLTEGDDDEVIKYNLSR
    LPANVEYLAFTVNSFRGOSFNDVENAFRCRVDDGTKEKLARKYKLEQSGSHGIVLSLR
    RNNGMDFALGHACRGRTIDMHSDIVSAVIR"
    /note="tellurium resistance, similar to TerA proteins e.g.
    [Serratia marcescens] gi|5702379|gb|AA047285.1|af168355_3
    percent identity 67 in 385 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance protein TerA"
    /protein_id="BAB34775.1"
    /db_xref="GI:13360812"
    /translation="MVPYQPRNDGCVSLVSEGOYTFVVALNRKLPVOKIAFTVT
    CDGQTVSGLRNLSIDVEGATGLVSGSVLSGQENALILGEFYRRNNMKEFFVQ
    GFNGGLPLAEHGVNIADPPAPAPAPPPVATPPVETKISLSVSLTKKPAISLTK
    RDNFGEIRINLMNHRGSSKSGFAGMEFGKSIDLDLGAFELODQKSVIOLAGNAFSD
    YRDEPYVOLGDRGTGSDGEMLHNGRBMKHIREVLIAFLIEGVPSMDKTDGYVT"
    /note="tellurite resistance, similar to TerB proteins e.g.
    [Serratia marcescens] gi|950680|gb|AA086848.1| percent
    identity 100 in 151 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance protein TerB"
    /protein_id="BAB34776.1"
    /db_xref="GI:13360813"
    /translation="MSFFDKVGKAINSGRDELTRQVGRFNKKFMQGTVAVCARIAS
    SDVSSSEKQKMGFLRSSSEELKPDFTNEVIEFFNKLVSFDPDVEIGKETMYILA
    LKQPEAOLALVGTAVAKSDGNFPQDEKLARETALNALGPPAERGL"
    /note="tellurium resistance, similar to TerC proteins e.g.
    [Serratia marcescens] gi|6226214|sp|Q52356|TERC_SERMA
    percent identity 100 in 346 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance protein TerC"

gene
CDS
    /protein_id="BAB34777.1"
    /db_xref="GI:13360814"
    /translation="MSTHIGFPIETVYVIALSVGAIFIDLFMHRDDKPISLKSAAL
    MSFVWVYVMAFAGFLYTHGEAVASLFTVYALKEKVSYNLFWYMAIFSPFVPPDR
    YRHLRYIGTIGATVFRGIIIPALIGKSLSTCPRYVEYFAITVAANTAYMLKSGDDDE
    IEDYSQHLATKMYKRFPIPWKLRGHAFLLQKEDEDALEPESDVTIGKKRAALY
    ATPLECAVAVLESDVMFEDSPVPAIIAVSEPLIYVSNAMFALGLKTLFVEALK
    QYVHLKNAVIVLLEFFIAAKGLNATDHIWMHGYSIAATSTLYVVLGVALGILASVW
    FPGKESEKSGS"
    /note="tellurium resistance, similar to terD proteins e.g.
    [Serratia marcescens] gi|6094448|sp|Q52357|TERD_SERMA
    percent identity 100 in 192 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance protein TerD"
    /protein_id="BAB34778.1"
    /db_xref="GI:13360815"
    /translation="MWSLSKGVNSLSKAPSMKNVLYGLGMDARSPDQDFPLDS
    AFLASNGKVRGSDDFRPNLSTSDGSYTHRGKRGREGDDESLKIKIDAVPEY
    DKITFYTIHQARROSGFQVSCAFIRLVNDQTEVARYDLTEDASTFAMLFGL
    YRHGEMAFRAVGGGVAGLASVCAQGYINS"
    /note="possible tellurium resistance, identical to
    gi|17108482|gb|AA036434.1|af126104.3 TLRB_ECOLI percent
    identity 100 in 191 aa, similar to TerE proteins e.g.
    [Serratia marcescens] gi|6094449|sp|Q52358|TERE_SERMA
    percent identity 98 in 191 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance TerE"
    /protein_id="BAB34779.1"
    /db_xref="GI:13360816"
    /translation="MAVSLVGVGVSILTKCAPNNVAMVGLMDARTDQGFPLDAS
    VENGEGKVLSDMHLFFFNKRTSPDCAVEHOGKNGRTGEGDDEQVKTITKVSADI
    KTLFAVITIDAEARKQNGFVNSFMRYVNNNDGTIARDLSEDASTAMVFGEL
    YRHGTEWKFRAVGGGFAGGLSALASQGVNV"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34780.1"
    /db_xref="GI:13360817"
    /translation="MGGFGLIDPLSTGEYKGGGRQSYVNSGMNLCRPVYVLL
    VKASIPCLISTACPSDICFCDL"
    /note="tellurium resistance, identical to
    gi|17108481|gb|AA036433.1|af126104.2 TRLA_ECOLI percent
    identity 100 in 102 aa, similar to TerF proteins e.g.
    [Serratia marcescens] gi|17387491|gb|AA086852.2|
    TERF_SERMA percent identity 94 in 102 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tellurium resistance protein TerF"
    /protein_id="BAB34781.1"
    /db_xref="GI:13360818"
```

| | | |
|------|--|--|
| gene | | /product="hypothetical protein" /protein_id="BAB34765.1" /db_xref="GI:13360802" /translation="MSKPRWTLDOCKKHVAAMRASGLTROYCELYDIPKSLRMPQ DVAKKARARAPELIVSVSGSSGMDGRSLDSEPTLFLRGGIRMKCCQSQGLDVR ALRHADA" complement(268292..268759) /gene="ECS1343" complement(268292..268759) /gene="ECS1343" /note="immunity to R478 phage/colicin/tellurite resistance cluster, similar to Terw [Plasmid R478] g111354147 gb AAC44736.1 percent identity 99 in 155 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="Terw protein" /protein_id="BAB34766.1" /db_xref="GI:13360803" /translation="MQLNTRQARIFKLANLGTGKPVSAADIITLSECSPTLRLAK ELRESYSAETIKYSKAGSHYHLVNPGLDKKTLRRMNAELQNAELKGESTGKVLDDK DKTAVSLSRMILRKIDRLAIDSTREAVENKALHSVDLIKREYSAKKS" complement(268777..269985) /gene="ECS1344" complement(268777..269985) /note="unknown" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34767.1" /db_xref="GI:13360804" /translation="MGFRPKSINIIPGVRLNSNGAPSLISVGRGASVFGSGRTYA NGLPBTGLSYRTLRDRAARKSGENRTAIDPGLQALBQKAELMSAVTAIRNIHET PDKRTGISVAELEVAVYLHNRTSPQVAPVRPEKPDVLVPEKPAESEGISFGKWE SESASKERHAEIRNWOQLDIVERENTLROHRYOQRTMAOQYANMKKEAEHNER LATAQADARQOFRDAAFEESYLAGVLAETEMPRETLVAEVAPBELSAVLDVDAI EDPPKITYGNARGETELTERKAMQKAVRENYAHNVGCLPRLGIVLHTLPFNVIYS GFMORVSKPTGYLEDEYILSCKTRSQMSVNFAGIKHIDPVALGDQPIRKMSSTF IFQPIRPLTL" complement(269996..270952) /gene="ECS1345" complement(269996..270952) /gene="ECS1345" /note="unknown, similar to hypothetical proteins e.g. [Deinococcus radiodurans] g117472167 pir B75302 percent identity 42 in 305 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34768.1" /db_xref="GI:13360805" /translation="MKNRLSPWNIGATLYMPATREDIADAVLHGKIPGLSLVLCED AVSEADIPALKNLLEHLHLSNSMSGLKNDPLVEIRPRHAEKMGWLKAHYDLSV DGFVLPRFTLSSIAEMWMDIMVGHILCMPTLETEDVEDVYQWRELAIRLEHCHDR IAYLRIGNDLNMVSLRRPRDLTLTDSKPMYTKMLVSYFGPPDFALTAIVYCHIDR IWAARELADMAHGLVGLKTAIHFGQIEVIONALMTVGESDRLRLNSTQAAVFKSOG AMCEPATHRRAAGILDRARFYGLQNEOSADGIRLRLTVQDH" complement(270952..271686) /gene="ECS1346" complement(270952..271686) /gene="ECS1346" /note="unknown, similar to hypothetical proteins e.g. [Streptomyces coelicolor A3(2)] g117472048 pir A75302 percent identity 34 in 260 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34769.1" /db_xref="GI:13360806" |
| gene | | /translation="MGKTSWHYGISIIRDRIGDAALDVEERHGTSGIVEVDGNTGK GAITGEVRLAKLDRPGYDEQRLVLAADPCGCSLAASDDMLIPFGIMGAPVSGLIS RSWSSGELHGCWCENHSEFECSSRLMADVAHFRKRLTPSSLAIPLSNTEASARYLW TSRVDVIFLADEFKVDVSNRIKPCIAETRAVLRLVRPDHVEVRSIDDDVVALVGLAR EKGIYVEMGCTLQGYRAVYTIKKYV" complement(271688..272068) /gene="ECS1347" complement(271688..272068) /gene="ECS1347" /note="unknown, similar to hypothetical proteins e.g. [Streptomyces coelicolor A3(2)] g118246803 emb CAB32838.1 percent identity 45 in 97 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34770.1" /db_xref="GI:13360807" /translation="MPHSEFLERIKMQPHETFGSYQPGDVEFLKRPVIENTPVEK KEELIQSGKHYSDMLSQEPAPTOQWHLDFRRLDRGAEIRLAKEVYQIALAERFGD EPYIASLVRAGVPLGRVAAGPA" complement(272034..272807) /gene="ECS1348" complement(272034..272807) /note="unknown, similar to hypothetical proteins e.g. [Serratia marcescens] g111695868 gb AAB37122.1 percent identity 100 in 167 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34771.1" /db_xref="GI:13360808" /translation="MKNPVVLSDDTLFQTRRKMDVDELAEPRTGAVDRTLNPFRF WKEQMSLVWMLQAEILPVTARGTEBIESVRIPFMSWATTTGAVLTTPREGPDER MTAEMLQLAPYOEKLSMORLITLTEMMDAGINMARNFETGMAVYMKHHDSER LDENALADELENYEFPPEGFYIHNSNVNVAWLPVPEKGLAVSWLLEKLRARGVFPY IGCDNSTDRHMKTCMPFGIPROSGPADNISQRIFFEN" complement(272800..273942) /gene="ECS1349" complement(272800..273942) /note="unknown, similar to hypothetical protein e.g. [Serratia marcescens] g111695869 gb AAB37123.1 percent identity 99 in 197 aa, [Deinococcus radiodurans (strain R1)] g117471591 pir B75301 percent identity 38 in 364" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34772.1" /db_xref="GI:13360809" /translation="MSDILEDMYIRRTLSCGTIOVTRDQGEVSLDDFLDIARRNPKR AFLEFSKVLGRHIEPVPSVKAQVROLASQPSLTGTPVLEITGMAETAVAGAGVDFE VRHQRHSVYLITSRHVDGTLCEFEKSHSHDHLIYLPDDEKRRRVNARTIVL IDDEATGNTFNILSLNRNKGKQHEVOYIATVLTTPMSNALESRIPLPYTSLSVS GQMGWTPLPDPAVPDMKRVNTSGENDTQGSWGRGLMAPADLGEHVSVKGER ILVIGCEFWPEFLAERLEAAGAQAFYSTGSPVAVGVAISAISFTDNYGLGIP NFVYNVAHQQFDRILVCTETPESIDPQLLKALAEVAVVEIYVE" complement(273952..275010) /gene="ECS1350" complement(273952..275010) /gene="ECS1350" /note="unknown, similar to hypothetical proteins e.g. [Serratia marcescens] g111695870 gb AAB37124.1 percent identity 99 in 173 aa, [Serratia marcescens] g111695871 gb AAB37125.1 percent identity 98 in 53 aa, [Deinococcus radiodurans] g117471522 pir B7530" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" |

| | | |
|------|---|------|
| gene | <p>/note="unknown"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34751.1"</p> <p>/db_xref="GI:13360798"</p> <p>/translation="MISLPSGTRILWVAGYDMKRSFNGEGEOVHLDENPFGHLEIFKRGVDITKILMDADGCLFTKRLLEGQFTWPAVRDGVKSITRSQMLMLDLDLMROPKTSRLNALTMWL"</p> <p>264818..266431</p> <p>/gene="ECs1339"</p> <p>264818..266431</p> <p>/note="unknown, similar to hypothetical proteins e.g. L0015 [Escherichia coli] g134148831gb AAC31494.11 percent identity 61 in 503 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34752.1"</p> <p>/db_xref="GI:13360799"</p> <p>/translation="MSRKILIRITTELERLLSQEAELRLKQDQLSLVEETAFRLSALAAEKEITDEERIEHLRLQIEKRLMLRFTGRREKIQREVEQEAOLKQEQSDRSYGREDDQVPRQKQSHNRRLPRLHRLPEECCPECESELDYLGVEADELTVSSALKVIRTVAKKCTKDCIVAPASRPIERGIAGSLARVLTKYQENPLKYSEIFARQGEISRALSNWDACCQMLTINDLIRYVMTRKRVHDDTPQKYLAPCKKAKTGIRIMTVYRDDNRNAGSSPPAWFAASPRQGNHPQHLRPFQIGIQADAFSCYDRFASNRGEGALTEYACMAHNRKTHDVTISSATAEALRISLVAIEDEIRKLPESERLAVRQQRKVLTLTSLHEMWKRNLTSKRSRGACSVYLNOMDALCVYSDDGLAEADNNAERALARACLGKKNFMFQSGDHSGERGALLYLGTCRLINDPEAYLRHISVLPEWPSNRVDELLPMNVLTNR"</p> <p>complement(266435..267163)</p> <p>/gene="ECs1340"</p> <p>complement(266435..267163)</p> <p>/note="unknown, similar to hypothetical proteins e.g. 57.8 kd protein [Pseudomonas putida] kd 24967401sp P556301Y4Q1_RH1SN percent identity 37 in 232 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34763.1"</p> <p>/db_xref="GI:13360800"</p> <p>/translation="MKLARQQRFGKCKESLAGNQRLFEEDVDADIAEISAHLDKLLPQTGDEKTTTRPVKRPPLPSPLPRAEKVIVIPAEERCPDCAPLHFIIDVESEKLEYIPAQVYVNRXYIRPOYSCPCKEVFSGMAKHLPLSAVPSYIAQVYISKYDHLPLEYQOHIIFSRMGVELPVSYTMADWVGVALAPLAKLLRHLLRLDVIHADETSRLDLTKGSKSCSGLMAYVYSGSERVSNCAFY"</p> <p>complement(267374..267736)</p> <p>/gene="ECs1341"</p> <p>complement(267374..267736)</p> <p>/note="unknown, similar to hypothetical proteins e.g. 20.3k protein [Agrobacterium tumefaciens IS1131] g11950901pit1 JC1151 percent identity 41 in 101 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34764.1"</p> <p>/db_xref="GI:13360801"</p> <p>/translation="MESGLFAMLTDPNNFLVVKPYDMRGIDTLTOYVONELNAAMHDGAAPFTINKVRSRIKVLRLMDKHGWLCTFRLHRGSRFMRKRDATWHLTQDEFHMLVFGVDQOQVKGCHDLAKWYO"</p> <p>complement(267702..268031)</p> <p>/gene="ECs1342"</p> <p>complement(267702..268031)</p> <p>/note="unknown"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> | gene |
| CDS | <p>/note="unknown, similar to hypothetical proteins e.g. orf50 [Escherichia coli] g1160094261db BAA8485.11 percent identity 71 in 106 aa, L0014 [Escherichia coli] g113288157 emb CA11510.11 percent identity 64 in 116 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> | CDS |
| gene | <p>/note="unknown"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34758.1"</p> <p>/db_xref="GI:13360795"</p> <p>/translation="MHTNDRSLTWDFLFCAGATARLVSEAGSEELRYCSEGLITLTHSGEPLAHTIDHRLTWDFLFCCTGATARLVSEAGSEELRYASGREVLTIVLWAGDEVOREMLRAGSGDSLSARVYALAHLSVQVLPYAGCYGLHEHLMAGTTHPEKYNPSSRSGSSSGHGRVRSVSQACSRPRLITICARFPAISPPANNAFRPMTTGAIASQILSLRGTAIPSLERKQSSSRFRFI"</p> <p>complement(263210..263809)</p> <p>/gene="ECs1336"</p> <p>complement(263210..263809)</p> <p>/note="unknown, similar to C-terminal parts of hypothetical proteins e.g. [Pseudomonas putida] g1129956331gb AAC98738.11 percent identity 40 in 200 aa, L0015 [Escherichia coli] g1134148831gb AAC31494.11 percent identity 39 in 200 aaGTG start"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34759.1"</p> <p>/db_xref="GI:13360796"</p> <p>/translation="MYKSLADNDPGINSACMSHARGFANLYKASREPPRAAMLRKIACIVRLEKTRERPERKIRQWRQRTSPRYNDLPWPEEQPCPPGKPLKALNYILNRDELSCILSDGAVPLDNNICERAIKRVVGRAMLFAGSLMAGNRAAQIMSLLETAKRNGLEPHMLTDTLRLPEMPEDRLLELLPLEGFTPSG"</p> <p>263929..266461</p> <p>/note="IS682-1"</p> <p>264015..264440</p> <p>/gene="ECs1337"</p> <p>264015..264440</p> <p>/note="unknown, similar to hypothetical proteins e.g. orf29 [Escherichia coli] g1160094051db BAA84864.11 percent identity 37 in 136 aa, L0013 [Escherichia coli] g1134148811gb AAC31492.11 percent identity 38 in 124 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/evidence=not_experimental</p> <p>/product="hypothetical protein"</p> <p>/protein_id="BAB34760.1"</p> <p>/db_xref="GI:13360797"</p> <p>/translation="MEOKILISEPRRSFNSNEFLQWKLASQCAVSARIAREHDINDNLFLKRLMQONERRISRLRPVTTSSQAGVELPVEITPDEQKPPMALPLPLSTPQSTVYASACVFEFHNGMTELPSELLTIVLIRELTGNGR"</p> <p>264437..264887</p> <p>/gene="ECs1338"</p> <p>264437..264887</p> <p>/note="unknown"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> | gene |
| CDS | <p>/note="unknown, similar to hypothetical proteins e.g. orf50 [Escherichia coli] g1160094261db BAA8485.11 percent identity 71 in 106 aa, L0014 [Escherichia coli] g113288157 emb CA11510.11 percent identity 64 in 116 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> | CDS |

```
gene
256914..257588
/gene="ECS1326"
CDS
256914..257588
/note="urease accessory protein, similar to UreF urease
accessory proteins e.g. [Klebsiella aerogenes]
g1137097|sp|P18318|ureF_K1EAE percent identity 79 in 224
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="urease accessory protein UreF"
/protein_id="BAB34749.1"
/db_xref="GI:13360786"
/translacion="MPTPEKRLRLMOLASNSLPYGVYSGWGLEMAVEAGVEDNSAF
EHMOLOMOESFPAVDLPILRLRYRACEADSPSAGRWYLLACRETRERDERNR
GAFTRLVDWOPDCPAEMRKLCQOSOLTMAMVLGRWQISVSDALSLGSIWESIV
MAGVRLVPYGQLAAQOLIMRLCARVANNMDSALATPDHAIGSATPLASIASAREHETQY
SRLEFRS"
257600..258217
/gene="ECS1327"
257600..258217
/note="urease accessory protein, similar to Ureg urease
accessory proteins e.g. [Klebsiella aerogenes]
g1137099|sp|P18319|ureG_K1EAE percent identity 90 in 205
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="urease accessory protein Ureg"
/protein_id="BAB34750.1"
/db_xref="GI:13360787"
/translacion="MNTIKQPLRVGVGPGVSGCKTALLALCKSMRPTNOLAVTNDI
YTRDORILTEAGALEAERIVGVETGGCPHTALREDSMLAAEALSEKFGNLELTF
VESGGNLSATFSPELADITIVYIDVAEGEKIPKGGPGITKSDFLVINKTDLAPYVG
ASLEVMERDTLNRGERPWFMTLKSGEGLONIAIFEEQGMGK"
complement(258265..258546)
/note="partial CDS, similar to C-terminal parts of
autosecreted serine protease e.g. Pet1enteroaggregative
Escherichia coli] g113095184|gb|AAC26634.1| percent
identity 92 in 79 aa"
complement(258473..258814)
/gene="ECS1328"
CDS
complement(258473..258814)
/note="unknown, similar to hypothetical proteins e.g. TnpJ
[Shigella flexneri] g115532468|gb|AAD44752.1|AF141323_23
percent identity 100 in 87 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34751.1"
/db_xref="GI:13360788"
/translacion="MSRKTORVYSEKPKAEAVPTVLENOQLSISGASRLSPEGLIGW
VTAARKGIGTPGSRVAAELESILQRLKALNEARLERDILKKAAYDSICCNIPFNIL
FHCPRRYRLSV"
258505..258551
/repeat_region
258505..258551
/note="IS30-2, truncated (5')"
complement(258552..258878)
/note="IS600-2, truncated (3')"
258853..259008
/gene="ECS1329"
258853..259008
/note="ECS1329"
/note="unknown, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34752.1"

misc_feature
/gene="ECS1328"
complement(258473..258814)
/note="unknown, similar to hypothetical proteins e.g. TnpJ
[Shigella flexneri] g115532468|gb|AAD44752.1|AF141323_23
percent identity 100 in 87 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34751.1"
/db_xref="GI:13360788"
/translacion="MSRKTORVYSEKPKAEAVPTVLENOQLSISGASRLSPEGLIGW
VTAARKGIGTPGSRVAAELESILQRLKALNEARLERDILKKAAYDSICCNIPFNIL
FHCPRRYRLSV"
258505..258551
/repeat_region
258505..258551
/note="IS30-2, truncated (5')"
complement(258552..258878)
/note="IS600-2, truncated (3')"
258853..259008
/gene="ECS1329"
258853..259008
/note="ECS1329"
/note="unknown, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34752.1"

gene
259429..259692
/gene="ECS1330"
259429..259692
/note="GTG start
ribosomal protein L31-like protein,
similar to hypothetical proteins e.g. ribosomal protein
L31 homolog ykgM in intF-each intergenic region
[Escherichia coli K-12] g113023204|sp|P713021|YKGM_ECOLI
percent identity 93 in 86 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ribosomal protein L31-like protein"
/protein_id="BAB34753.1"
/db_xref="GI:13360790"
/translacion="MKPNIHPEYRTVVPHDTSIDYFRIGSTIKTRVIELDGVTPY
VTIDVSSKHPFYTGKLTETVASEGVARFTQRFGRFVAKKG"
complement(260066..260347)
/gene="ECS1331"
complement(260066..260347)
/note="unknown, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34754.1"
/db_xref="GI:13360791"
/translacion="MPTLTSSAPFYGAFCAPITLLPHFRDANKKASHFYTADFCFF
VICVGGDFRRLWQPDGSCPCRCGLRFTLQGRNRNRPHEAYDEFNR"
complement(260619..260960)
/gene="ECS1332"
complement(260619..260960)
/note="probable colicin immunity protein, similar to
colicin immunity proteins e.g. colicin E1 immunity protein
g111243951|sp|P029851|IMM1_ECOLI percent identity 25 in 107
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative colicin immunity protein"
/protein_id="BAB34755.1"
/db_xref="GI:13360792"
/translacion="MNRKFFKNTLMGNCPIITISILHLIDTPPEASIKVDLFLSG
LALPCSIYLTFTITLKYSISFMWPKPLPFWSGIYHFHLVYIMICIVLAIPSAIIVLFF
HFVGNITRNMK"
complement(261006..261677)
/gene="ECS1333"
complement(261006..261677)
/note="unknown, weakly similar to hypothetical protein
MAL4P2_26 [Plasmodium falciparum]
g116562728|emb|CAB62867.1| percent identity 29 in 106 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative membrane protein"
/protein_id="BAB34756.1"
/db_xref="GI:13360793"
/translacion="MEFSVIFFWVITGLITLFLKSIITDLIKNLKNTDIALYIAFL
SAGYTFQYINDNIFSELGAVGLFVTLIVINLPEFRLFLISKMLMSRHRKMQ
KKLKQTSVPDEIYYSIEKHKKYPIYVSPNNIIEIGYNIINDLVGGEKIHPRRTS
NRSILYENLTHPLPATIKDVFCTLHDYFYNNNNFTKSIDVYISKIKONHNPPLANN
ENVOE"
complement(261880..262035)
/gene="ECS1334"
complement(261880..262035)
/gene="ECS1334"
```

```
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative outer membrane protein"
/protein_id="BAB34740.1"
/db_xref="GI:13360777"
/translation="MTSYFNMSVVLWFLFTGILPCGLLLVNIIRYPETWIKGIYRLL
SMFASLLIFAIAPFYKDYASVGRNNSLKEKELIPNYYISGPKYVRDPFVSQEP
OTGDASRTINEKOKPYIMPLVGETARSONYALNGYSRGNDFTKYNELISHNNO
SCGSTRASVPCMSDMKRKEFNRSKAVNSENVLDILRTGVNLMLENDGGCGVCK
RIPTINIPESNDNTLCRKNSCYDEVMKNIDEYINNSEDKLIYFHLMSGHGPTYYL
RYPESHRYFKPTCDRSDIENCHEQLINTYDNTIRYTDYIISKLDLIEKDKEDYV
LLVYSDGESLGEGLYLGHTPYWVAPAEQTHVPLITWMSGFYSSKKIDLNCSEHA
LNRVSDNIFSSLGLGLMNVNTSVYKNEDDISSCR"
252369..252512
/gene="ECs1318"
252369..252512
/gene="ECs1318"
/note="unknown,TTG start"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34741.1"
/db_xref="GI:13360778"
/translation="MNVAKTTSIINKRKLIFLLIPPLIRATNCPKVIYFY
KTV"
complement(252667..252942)
/gene="ECs1319"
complement(252667..252942)
/gene="ECs1319"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34742.1"
/db_xref="GI:13360779"
/translation="MPEHSGHADNDNPFPSMGSDSIFMHCLOSYGKSTCDYVHTFVA
DVAPESLTGLQTYLYLTAGLLCDIVFHEGGPPMNGIILFPDIIIP"
complement(252994..253197)
/gene="ECs1320"
complement(252994..253197)
/gene="ECs1320"
/note="unknown,GTG start"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34743.1"
/db_xref="GI:13360780"
/translation="MSFRLLNNTVCFRHEAFRLPLGLYIARKAMHIVGIIQRIQKDP
MGCDYAIWLMRMISRCYGDNLIIQ"
253278..254021
/gene="ECs1321"
253278..254021
/gene="ECs1321"
/note="urease accessory protein Ured, similar to Ured
urease-associated proteins e.g. [Klebsiella aerogenes]
g1137310781sp|Q09063|URED_KLEAE percent identity 71 in 242
aa,TTG start"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="urease-associated protein Ured"
/protein_id="BAB34744.1"
/db_xref="GI:13360781"
/translation="MLSTALPKNKGMQAMLDLRFORLHGKTLITRRHVGILTYORP
FYPEEETHCHLILPPGGIVGDELIINATIDSDCHTILITPGASKYRSGCAOHO
ONTLTCNNATLEWLPDSDIFPFGAHATLHVPHLSSSTILANDLCLGRPVIGETFS
HGTLANREIWDGSPULIERLHVANGELFCVARKRWGTMPLFLDNETQLEDIREKL
TPLENAGATLIDGLTVRFLSDNLRCQ"
254112..254414
gene
CDS

/gene="ECs1322"
254112..254414
/gene="ECs1322"
/note="urease gamma subunit, similar to urease gamma
subunits e.g. [Klebsiella pneumoniae]
g11370841sp|P18316|URE3_KLEAE percent identity 96 in 100
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="urease gamma subunit"
/protein_id="BAB34745.1"
/db_xref="GI:13360782"
/translation="MELTPREKDKLLFTAAALAERLARGLKINYPESSVALISAFIM
EGARDGKSVAAIMLMEGRHVLSREQVMGIPPMIDIOEATFPDGSKLTVYHNPII"
254423..254743
/gene="ECs1323"
254423..254743
/gene="ECs1323"
/note="urease beta subunit, similar to urease beta subunits
e.g. [Klebsiella pneumoniae]
g11370771sp|P18315|URE2_KLEAE percent identity 82 in 106
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="urease beta subunit"
/protein_id="BAB34746.1"
/db_xref="GI:13360783"
/translation="MIGEYVKRQYIEINGRATCSIIYENHGRPIQGSHTFHE
VNPALKFDRQKARQYRLINAGTAIVRREPQGRVELVALSGARIVHGFQDINGELE
ANDE"
254733..256439
/gene="ECs1324"
254733..256439
/gene="ECs1324"
/note="urease alpha subunit, similar to urease alpha
subunits e.g. [Klebsiella pneumoniae]
g11370701sp|P18314|URE1_KLEAE percent identity 90 in 567
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="urease alpha subunit"
/protein_id="BAB34747.1"
/db_xref="GI:13360784"
/translation="MMSNISROAYADMEGPTGDKIRLADTEMLIEVEDLTYGEEV
KFGGKVIIRDMGCGMLSAGCAADLYVTNLIIDYGVYADIGCKRGRIFAIGKRN
PDIOPNVTIPIGVSTETIABEGRTVTRGVDTIHIMTCPODAEALTSGLITMIGGT
GPTAGSNATCTCPGWTIVOMLOADSLPVNIGLLGKNGSNPDALKEQVAAAGVIGLK
IHEDWGATPVAVINCALIVADEMDVVALHSDTINSGFVEDTLITLIGRTIHFTEEG
AGGSHADITIAACAHNPIILPSTVPTLPYVNTIDEHLDMVMCHLIDPIADVAFA
ESRIROETIAEDVLDHDLGAFSLTSSQOAMRGVEVLRTWVDAHBMKYORBDLPPE
SGNDNVKRVRYIAKYINPALTGHGIAHEGSIENGCLADLVIMSPAFKVRKATYK
GGMITAMPKGDINGSIPTPOPVHYRRPFAALGSRRCRTFPLSQAAAAGVABQULM
HSTTAVVKGCRIVQKADMRHNSLLPDIIVDSQTYEVRINELITSEPADILPMAORF
LF"
256449..256913
/gene="ECs1325"
256449..256913
/gene="ECs1325"
/note="urease accessory protein, similar to Uree urease
accessory proteins e.g. [Klebsiella aerogenes]
g11370951sp|P18317|UREE_KLEAE percent identity 80 in 154
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="urease accessory protein Uree"
/protein_id="BAB34748.1"
/db_xref="GI:13360785"
/translation="MLYTRRVEFPAGQTTASVTLVPDMRVKSRITKYTLNDGROAGILL
PRGLLRDGDILSENDEIRIKYIADENAVSVRCADPFMLAKACHLGNRHPDLOIM
```

gene

```

gene
CDS
/translation="MIPLPSCGKIMLVAGITDMRNGFNGLAAKQVQLKDDPSGHAF
IPRGSSQVALLMSTGDCGLTKRLERGNFAPPSARDGKVPFLTQQLMLLEGIDM
RQPKRLTSLTML"
complement(246695..247096)
/gene="ECSi309"
complement(246695..247096)
/gene="ECSi309"
/note="unknown, similar to hypothetical proteins e.g. L0013
|Escherichia coli| g113414881|gb|AAC31492.1| percent
identity 98 in 133 aa, GTG start"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34732.1"
/db_xref="GI:13360769"
/translation="MGTVSDMKQNVTPGRRKCGPNYPPEFKOOLVAASCEPGISISK
LAEVGNVIANLTFKRWQRMREGKLTLLPSSESPOLLPTVLDAARQPSLADEPEPLSI
SCVETFRGCTLRFGNVSEKLTLLIQLQEK"
complement(247151..247462)
/gene="ECSi310"
complement(247151..247462)
/note="IS30 transposase (interrupted), similar to
N-terminal part of IS30 transposase
g12851554|sp|P372461|TRAB_ECOLI percent identity 99 in 101
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="IS30 transposase"
/protein_id="BAB34733.1"
/db_xref="GI:13360770"
/translation="MRRFTFAEERKASYFELMKNGTFSEIANILGSKPGTITPMLRDT
GAIKHEKRAVAHLLTSEEREIRAGLSAKKSIRAIATALNRSPSTISREVQRNRKRT
A"
247523..248463
/note="IS600-1, truncated (5')"
247602..248420
/gene="ECSi311"
247602..248420
/gene="ECSi311"
/note="probable transposase, similar to transposases e.g.
Hpi1 |Escherichia coli| g113614821|gb|AAC61713.1| percent
identity 98 in 272 aa, insb [Shigella dysenteriae]
g155332671|gb|AAD44751.1|AF141323_22 percent identity 98
in 272 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transposase"
/protein_id="BAB34734.1"
/db_xref="GI:13360771"
/translation="MCOVEGYSRSGYVNMVQHEPSDRKOSPERIKLEIKVAHRTRET
YGTIRLQLELAENGITVGRDLARKRLRKRCOKKRFRTTNSNNILVYAPRLNO
TFAPAPQWVWVADLTVAIQEGWLYLAGIKDYVTCIEVYAGMERMTKLTGKALEM
ALRSRPPAGGIIHSDRGSQYCAVDYVIODFGLKTSMKKGNADAPESWEGTLL
KNESLSHPFRKSDPDAISIVIREYIEIFYNQRRSRLSGNISPAFIREKHQMA"
248326..249072
/gene="ECSi312"
248326..249072
/gene="ECSi312"
/note="probable complement resistance protein
precursor, similar to lipoprotein trar precursors e.g.
g1418135|sp|P32885|TRTL_ECOLI percent identity 83 in 227
aa, TTG start"
/codon_start=1
/transl_table=1

```

/evidence=not experimental
/product="putative complement resistance protein precursor"
protein_id="BAB34735.1"
/db_xref="GI:13360772"
translation="MRFSIVSVVTLWVGISPPROPSGKNIIRMLIKRPTNCSVRYSQCO
NSTIILPEASERTVFLOIKNTSDKMSGLDOKIDAAVKARQYQIMTSPDKAYVVOAQA
VIKAHKMDLRSGQSLSRGSEGAANGALGGGITPITRYNGSAGTTIGVLAVLGLGMAR
DAWVEDINTMTVDVAIERFTVSRTINVALRGVISGSKIQTSTEGNOHKTOTFR
VSSAOVNKLFEPAAPHELDOLAKSIAMIL"
complement(249120..249371)
/gene="Ecs1313"
complement(249120..249371)
/gene="Ecs1313"
note="unknown"
codon_start=1
transl_table=11
evidence=not experimental
product="hypothetical protein"
protein_id="BAB34736.1"
/db_xref="GI:13360773"
translation="MESEFODSPFPYSGDLVYLREKVELTEELKYCSCIRNRKHKKY
SYGRQANRIKLNLVPLSDSVPAWYGVTGKIISELSER"
complement(249395..249685)
/gene="Ecs1314"
complement(249395..249685)
/gene="Ecs1314"
note="unknown"
codon_start=1
transl_table=11
evidence=not experimental
product="hypothetical protein"
protein_id="BAB34737.1"
/db_xref="GI:13360774"
translation="MNLSSAMDAGDEGELLMSGNMOPFNICRRDNHYPERKLIPLLE
ELFDNYGIELNKLETKEDSGSINFVSTSKNNQSASAVRLKDGISPCQGIC"
complement(249761..249961)
/gene="Ecs1315"
complement(249761..249961)
/gene="Ecs1315"
note="unknown"
codon_start=1
transl_table=11
evidence=not experimental
product="hypothetical protein"
protein_id="BAB34738.1"
/db_xref="GI:13360775"
translation="MIDPADISRYAQQLMIQGFISARYSQIDDLFAQLMFVLSQPGFC
GHHSQPMGNRRRIAAVMRFCR"
250371..250730
/gene="Ecs1316"
250371..250730
/gene="Ecs1316"
note="probable diacylglycerol kinase similar to
diacylglycerol kinases e.g. glj12521|sp|P00556|KDEL_ECOLI
percent identity 76 in 119 aa"
codon_start=1
transl_table=11
evidence=not experimental
product="putative diacylglycerol kinase"
protein_id="BAB34739.1"
/db_xref="GI:13360776"
translation="MANTVGTVRTIKAGYSWMKGLCAAMRYEAFOECIIIVLSALLI
ISFWIDVDSRVLLIGSAMVYIIEINSATLEAVVDRIGSBHBLSGRADKMSAAMV
LTIIIVAIIMGVILLIS"
complement(250823..252109)
/gene="Ecs1317"
complement(250823..252109)
/gene="Ecs1317"
note="unknown, similar to hypothetical proteins e.g. yjdb
in bass-ady intergenic region [Escherichia coli]
glj171986|sp|P30845|yJDB_ECOLI percent identity 45 in 428
aa"

gene
CDS
/protein_id="BAB34724.1"
/db_xref="GI:13360761"
/translation="MALICELDEQMSFVENKARQGMHWYAYKTKADGVLAATFEGPRD
ETCRLEPFLKPFSGAMITRONRSSYTEMPODKHLVGKIFTRIERNNLTLRTHIKR
PARKTICFLRSLEIHEKPLVHLSKTHVLLTGVTFRASFAVFLP"
237783..238079
/gene="ECS1302"
237783..238079
/note="ECS1302"
/note="probable regulatory protein, similar to prophage
cp4-57 regulatory protein Alpha [Escherichia coli (strain
K-12)] gi1461502|sp|P33997|ALPA_ECOLI percent identity 52
in 61 aa"
/codon_start=1
/transl_table=1
/evidence="not_experimental"
/product="putative regulatory protein"
/protein_id="BAB34725.1"
/db_xref="GI:13360762"
/translation="MPDQSGCRTPSRILRMPELSQLIGISRTIYEKMPLSKYDA
TFPPRVLGSSGVGRSSAIDEMLTLLHTVPARSAKGVNDEVRRHTLSSDWTMSR"
238722..240107
/gene="ECS1303"
238722..240107
/note="ECS1303"
/note="unknown, similar to hypothetical protein b2625
(YfiJ) [Escherichia coli K-12]
gi1173621|sp|P52124|YfiJ_ECOLI percent identity 40 in 444
aa"
/codon_start=1
/transl_table=1
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="BAB34726.1"
/db_xref="GI:13360763"
/translation="MCLAPENPYPIYALPPLVNRNATITETOKNQPAPLAVATSA
LTA ISACNOIDVCRGNLHGPVNLSTLADSGERTYDVKPMKAFYRDALDEVA
KLIVENSTKEIKMEQKOKALESPKHEKTRACKDKVATSELETILNRPVPPQIRRTI
FNERTIEGMILKYSDNSRSPALVSEGGVITDSRMSKLGITITLMDGSLFIDKSS
PGINLKEPRLTISWIDPDIYHKGFCYRKKEIVTSGHARFLMCOPTSTGTRIMN
ELIDESLAMSGERCLHSPQARIMTDYNDVESKLGGLPLRHCRAYAKNEYMA
RLAGLIIHSSGEEGELSPYTAEMARELAIVNGNEVRLSNPLTFDMSALTVPRLIPE
ELLEFNMIKSYCIEKIGLICMKNKNDLDRGPRFRKKDKINLDDLTYEQNVVYIEC
KTLCAVNPNDL"
240928..241491
/gene="ECS1304"
240928..241491
/note="ECS1304"
/note="unknown, weakly similar to hypothetical protein
Cj1244 [Campylobacter jejuni] gi1696677|emb|CAB73498.1|
percent identity 25 in 78 aa"
/codon_start=1
/transl_table=1
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="BAB34727.1"
/db_xref="GI:13360764"
/translation="MYNANPYEMDMILKDVNHEMGLQRFSEKLLPFRIDFAVRKD
TPPSGHSRSMCIEMHRLSETQTMLAGYVMEYSNKGILHIFIGYLDGGRHNS
YRISROGLDMRRTIEGEGYFLHCRANKDYPVRIIDVYIHSKSAVDLRLALSYLAK
QDOKEHGIIILRSRLPEKSNRGRPRHN"
241646..244006
/gene="ECS1305"
241646..244006
/note="ECS1305"
/note="unknown, similar to hypothetical protein A153R
[Chlorella virus PBCV-1] gi17461298|pir|T17644 percent
identity 32 in 365 aa, also similar to DNA repair protein
rad5 PAB0128 [Pyrococcus abyssi (strain orsay)]
gi17514780|pir|A75209 percent identity 28"
/codon_start=1
/transl_table=1
/evidence="not_experimental"

gene
CDS
/product="hypothetical protein"
/protein_id="BAB34728.1"
/db_xref="GI:13360765"
/translation="MWHKSDSEDLAALRAENRATLSLEAHGIERRRKPPQPVORVS
LSTDEKVALFRLEFRGDRDVALRWESTSGSSGSPACANMOKARICGKAPKRGDC
AHROLIPYSDIYVYHHLACTHTVGMVPLLEDSCFLFVAVDDEAKQDASAFMNSCD
ELGVPALEISRSROGAHWIFFAVSARVREARLGTI11STYCSRTQRLRGSYDRL
FPMQDTMPRGGFNLIALPLQRRPELGSVFDVMDLPYDPMALVYSVPMVODI
EPTILRATGSHIPDLVNFINEEDLGTPMWEKSSGNRLNI5IAEPLKTLTLANOYFEK
AQLPOLINRLIRLAFNPPEFYKAQAMRSYMNPRVIGCAENYPOHIALPRGLDS
VLSFLRDNNIAEFLIDKRFAGTECAVMGNLRAEOEAVSALIRPDGVCAPAFAG
KTVTAAYIARKVYRLLVHRTTELLKQNEPLAFIDAGBSTGICGGKAPPCGND
IAYVOSISROGEVPELVNRIGQIIVDECHHIGAVSFA11LAKETNARYLLGLTAPYIR
DGLPIIFMYCGALHRTAVRKPESHNLNLEVLIRSRFTSGHLPSDARIQDIFREIALDH
DRTVAIAEAMKAFEGCRKRVLTERTDHLDEIAVMSKLSPEILHGRSKRRAM
LISGNALPDPSPRILSTGRLIGGFPHPIDTLILAMPYSWKGTLOOYAGRLREH
TGSQVRILDPDRTYVPLLRMRMDKRGQGYKAMGRIIADDESYI"
244276..244428
/gene="ECS1306"
244276..244428
/note="ECS1306"
/note="unknown, TTG start"
/codon_start=1
/transl_table=1
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="BAB34729.1"
/db_xref="GI:13360766"
/translation="MHNQFPEIQPWVGLRNSQNCFHVIYSGICTARGMWYGNKCVF
LKITDI"
complement(244341..244724)
/note="partial CDS, identical to C-terminal part of IS30
transposase [Escherichia coli] gi11552823|gb|IAB08675.1|
percent identity 100 in 81 aa"
complement(244481..247554)
/note="IS30-1, truncated (5') and inserted by ISEc8-03"
complement(244726..247167)
/note="ISEc8-03"
complement(244763..246301)
/gene="ECS1307"
complement(244763..246301)
/note="ECS1307"
/gene="ECS1307"
complement(244763..246301)
/note="ECS1307"
/note="unknown, identical to L0015 [Escherichia coli]
gi13414831|gb|IAC31494.1| percent identity 100 in 512
aa, also similar to hypothetical proteins e.g. [Escherichia
coli] gi13288156|emb|CA11509.1| percent identity 99 in
411 aa"
/codon_start=1
/transl_table=1
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="BAB34730.1"
/db_xref="GI:13360767"
/translation="WNDSDDIFILKORLAEOELHALOEKSNRERETDHLQAO
LDKLRMNGSRSEKVSRIQAEADLNLQKESDTLTGRVDPVAPVRLQTRRKPC
PESLPREKRLIPAPCCPNCGSLSTLGEDTALDEMSAFVIRTVRRKNACTOC
DAIVQAPASPRIRGIAAGPLARLVLTSAEHTPLYROSEIYGRGCVELRSLDLSG
WVADACRLSPLEELAGHYVTDGKLHADDPVOVLLIPGNKTKTGRIMAVVRDRNA
GSAIADHVAIRPSALTEBALQIQTALTEADIKGMPAEOQLAROKRTPKLSLES
WLREKMTYLSRHSLELAKFAVALNQWPAITYANDGWVEIDNNIAENALRAVSLGRN
FLFEGSDHGGEGALVSLTGCTKLNVDPESYLRHLYGIADWPVNRVSELLPMRIA
LPAE"
complement(246351..246698)
/gene="ECS1308"
complement(246351..246698)
/note="ECS1308"
/note="unknown, identical to L0014 [Escherichia coli]
gi13288157|emb|CA11510.1| percent identity 100 in 115
aa, similar to hypothetical proteins e.g. orf50
[Escherichia coli] gi160094261|db|IABA84885.1| percent
identity 76 in 107 aa"
/codon_start=1

[illegible]

```

gene
220022..220303
/gene="Ecs1287"
CDS
220022..220303
/gene="Ecs1287"
/feature="probable acyl carrier protein, similar to acyl
carrier proteins e.g. Acpc [Streptococcus agalactiae]
g11486773|gb|AAD32036.1|AF093787_4 percent identity 38 in
86 aa"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="putative acyl carrier protein"
/protein_id="BAB34710.1"
/db_xref="GI:13360747"
/translation="MSTMDAIEQRKVLMTITETIQRINIORDEAIDDDTALFQNG
GFKLDSVDATETIVLLDKVFDIRIKISESDPSYMSINSLATFVIOKKND"
220339..221499
/gene="Ecs1288"
CDS
220339..221499
/gene="Ecs1288"
/feature="probable aminomethyl transferase, similar to
aminomethyl transferases e.g. g117450600|pir||C75088
percent identity 26 in 333 aa"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="putative aminomethyl transferase"
/protein_id="BAB34711.1"
/db_xref="GI:13360748"
/translation="MKSIAEFHLKNNAMGVNRRRLPSSYHDAMPYKAVRENALIV
DYSHLSIVSGDDAMALINQIVSADYSIIRDDQATISLVINEGTRIGGVYICSD
GYLLSEDIISAEFLIASMNTLILKAEELDIOSMPETODMRENNKAILLEGPISWEIM
SEIHGEDVITLPTRYEYVNTFEDLLFRCGKNGEYATVITGEOKLAQWKLLTVGCK
YLMQGTGLDQKIVRLNCPMDASLWEGQAVNVPVQLOMQAVQYDKDDFGKAVDEL
SOETGNKILGIMQAECEGIEAGDRVLYEGQGVYKALFSPALQRFYALTLERD
YAMSIDSGEYICVANGHIIIPAQSKCMPIYMLSMVSPTEHSYIDASKNKSGAA"
221573..224131
/gene="Ecs1289"
CDS
221573..224131
/gene="Ecs1289"
/feature="probable 3-oxoacyl-[acyl-carrier-protein]
synthase, N-terminal half is similar to
3-oxoacyl-[acyl-carrier-protein] synthase (EC 2.3.1.41)
[Bacillus subtilis] g117433750|pir||G69842, percent
identity 37 in 393 aa, C-terminal half is also similar to
g1174"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="putative 3-oxoacyl-[acyl-carrier-protein]
synthase synthase"
/protein_id="BAB34712.1"
/db_xref="GI:13360749"
/translation="MKTNRKRVITGLILSLAENIDQFREALNKKNGKDSERPA
KFEENARAELVLDIEYSDLNEDVIDSLNMAALMAKVGKDALSOAGINDSPTLKEP
GIMVGVSAAGTEAFLEFLFQRMEDPSLRKALPSCASSCCSSYSTILGIDGQELVAT
ACTASPNAYGMADYIQNGKSKTMLAVGTETLPTFAGFYALNVHMPDTCPTPFSQS
GMSIGGAGALVLEETEAHVARGATVIGELISVATSCDAFHETGPDPRASGAOVANK
AMANAGVTPEQIDYVNAHGTGEANDRIELAMKRVFANHDKLVSSTKSEFHNIGA
AGIVELIACIATLPDNRVILPTLNFSHARPNCDDYVYNEFRDRKVNIMKNNYAFSGN
NCCMILSMEPASIPVSTDSKRAVLSGVGAVSAIHTLNQVLENTIAOQHNVHIGSVT
PPDDTLEFAKKELLAVIDESNOFAELDEPEXSEATLPTEBSNKRFPQVSGLEPRHL
RRDOKKATRGCFALITALEBOAKRKIKRGDELGMIMGMSRKPQETVYKLOSL
KDPDKRVATSEFPGLSMNAIPTECGISEGIKGVTTLATGENSALGALTYGEIVON
LOPQVYLVGADEFPMSLYMDAVTOKIMHTAASDYQIYKDPQGVGEGACMILL
EDPOLVARGAEALVEVGVKSCSNYTFDASILEEKSSMALAITERALTHAGISGT
IDLVCGTSGNSDHSRIEIDATYNTFPAONPVPVNVNACGCFVASAGLNLTAIYT
DCIKQAVPAINTVTFEFDERNVFNQSKLALNHYLLVGAITEGNTYAFVILKG"
224379..225359
/gene="Ecs1290"
CDS
224379..225359
/gene="Ecs1290"
/feature="unknown, weakly similar to parts of polyketide

```

```

gene
227668..228978
/gene="Ecs1294"
CDS
227668..228978
/gene="Ecs1294"
/feature="unknown, similar to hypothetical proteins e.g. BbJ27
[lyme disease spirochete plasmid J/1p38]
g117463605|pir||D70248 percent identity 25 in 399 aa"
synthases e.g. [Streptomyces sp. strain MA6548]
g117481905|pir||117428 percent identity 23 in 201 aa"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="hypothetical protein"
/protein_id="BAB34713.1"
/db_xref="GI:13360750"
/translation="MLDKHCLAGRNRYVLLTGIPRIDGLDKSFYVNDVEDINLQ
SITNLHASKMSODTISHLARKRYLOYLPPMNCNTSSILAVILANQVIFQGGIDLA
IVLNCSTIKQDIIWFLEFQSLDSEQVQPRENSKQVAFAGFALLLESNNHRRKQ
QSEGVRYQTTQYIISAGRSNDASWLSTVNHVQMAKQAEIALDDLAAILPHNGSA
VSDNAEKALAMFAGEQIPLYAKGOIGYATGSGVVDLIIGHNSLTHHQLIAPVN
DVIIDSMASVILITDGSYVTHNSKRHLKLVYGVGSDSVIGVWNTNQAQRAK"
225458..226309
/gene="Ecs1291"
CDS
225458..226309
/gene="Ecs1291"
/feature="unknown"
/feature="unknown"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="hypothetical protein"
/protein_id="BAB34714.1"
/db_xref="GI:13360751"
/translation="MYERGLYGRANGMLKRFGEVNDIYAPERIRNPQYPLEAME
QVNDLSAQOMPTKRMKLMKSRPAFYADSGESGESEKISSALHITITLPKNLTK
KFSITTEPTKIRGEKQAFVOALRMAQOYLAQVNDVCAVAPVAVLPVSESDVCK
PKRLTPRRADDEVLTSEYRGCFILSDRESNIRQCEHYHSGRTPQATIDIVMASE
RQOMPSPRAGVIDLVAHYBEGCLTFLASMDYILTRHIOPEGKMRVIAQKQFYHF
DLETLG"
226320..227027
/gene="Ecs1292"
CDS
226320..227027
/gene="Ecs1292"
/feature="probable ABC transporter, similar to putative ABC
transporters (ATP binding proteins) e.g. [thermotoga
maritima] g117445988|pir||H72342 percent identity 50 in
222 aa"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="putative ATP binding protein of ABC transporter"
/protein_id="BAB34715.1"
/db_xref="GI:13360752"
/translation="MSTIKSLPAISLSUNIKCYGAGAGKVDALKSINLDINKGEFLA
LCGPGSGKSTLINLISGIDKPTSGFYIFLNKILNLOPEOVLAVIRKHHGTFQPN
LMPVLVNFQVVPVLYNGCFPSKKEATERTLHDSGLSGFDGRKFGSLGQOQAV
AIRALAHPEMVVIADEPTGNLDLATEAILEDLLITINDQGTTFVISTHSELKARA
RRVEIQDGVLIHDSDA"
227011..227667
/gene="Ecs1293"
CDS
227011..227667
/gene="Ecs1293"
/feature="unknown"
/feature="unknown"
/codon_start=1
/transl_table=1
/evidence=not experimental
/product="hypothetical protein"
/protein_id="BAB34716.1"
/db_xref="GI:13360753"
/translation="MTQMHKGFLLIGVLYAAFAVKRYAAPPAANOVLQOGLPAOP
AALATLTALKEKEVEARHNVALVFAAYGLRQADGAAATNDFIHASEVAKSGFPMID
EAVDLEKNOYRVLEARDAVAYPADSGRCVVYVKDEHMLADPAIMPTTIRPHIILAM
RYALALHCKDVTTRANLLNQIKQNMALADSLTFQDENNVPEMDSSEETQVLLPLMKG"
227668..228978
/gene="Ecs1294"
CDS
227668..228978
/gene="Ecs1294"
/feature="unknown, similar to hypothetical proteins e.g. BbJ27
[lyme disease spirochete plasmid J/1p38]
g117463605|pir||D70248 percent identity 25 in 399 aa"

```

gi|4105989|gb|AAD02646.1| percent identity 32 in"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative major pilin protein"
/protein_id="BAB34703.1"
/db_xref="GI:13360740"
/translation="MFSTFKKALALAIAPFSTMAAPVTFQGEVTDQTSVNIQO
TNSVLMPTVAMADFGATLADQAGOTPEPTVSVCQAPTGADQAIINTFELGYDA
TSGVMGNBDTSPDAKFGIQLMDSSSTGMPVLGATNVPGILTKVGTSEASYDFGA
RKFVDSAAATAGKITAVAEYTLSTL"
complement(210755..211378)
/gene="ECS1281"
complement(210755..211378)
/gene="ECS1281"
complement(210755..211378)
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34704.1"
/db_xref="GI:13360741"
/translation="MNKLLFRVITISDDLYVLOGISAKLGHHPFIECFCSQAC
NNDLSLRRDNRFTKVLVAVDVSTLAKPLARPLMIVSRKKMNRNIFSLSDQ
FTISRILKLTLLIKSTDKQVROVLTREMDIYLYEIKDKRILHTLLKRDY
KTIASHKISVFNFKEIACDSDGFYIKKALIQLRFTGFKVRLYQLH"
212775..216581
/gene="ECS1282"
212775..216581
/gene="ECS1282"
212775..216581
/note="probable filamentous hemagglutinin-like
protein, similar to hemagglutinin/hemolysin-related
proteins [Neisseria meningitidis] e.g.
gi|17225719|gb|AA04927.1| percent identity 25 in 1001
aa, also similar to filamentous hemagglutinin B precursor
[borderet]"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hemagglutinin/hemolysin-related protein"
/protein_id="BAB34705.1"
/db_xref="GI:13360742"
/translation="MINLSKEATVGRKALPPIALIMLSEPVASQAAGLYIKNGTVYNA
NGVPVVDINKPNSGSLSHNIMWNLNDKNGVFNNSANESSTSLAGNIGNSMLTSGS
AKVILNEVTSKNPSTINGMEVAGADKADLIANPNCITVNGSGSINTGKLTITGPD
IODDKIAGSVNGTITGKLDNAISPTETLSRVNVNGKVSADLVNNGVNNVYNAQ
KYNLSNSAQIOSASTINLTGTDNTTGTYSVGTISLNTKNTIVTRAGNISTM
GDIYVNSGITDNTNCKLAAGMLADPTNATLINSKGSSVGEAGLVALKGTGLTIS
NGQIRGVGLESALNNNGDIQTGDIATISNGVDNNKGLIRSTGHIYVGAAGS
VNNGSTKTADTSSDGLIADTGVETIGANNINNGGQIASNGVSLSSSTIDVAG
KILNSKVIITKSSLRNDGTGSGKQIEVAGKTLNNIGVSSSEGGISILANVD
NNGGFMGONTIMESGYNNTALIVASKIKLAKRSIENBDGNPFGAIGLFGM
POQTGVMGKEGIELSGQIVYNNNSRLAEDPVLQAQNTEDNTALVTSQADASIO
VGGTYNNATTVWSAGNDIDATTLQNSSSGTMDNNAAGFLASDKNLSELYVNSLTN
YGMISGKGVDVTVNNGNLNNTTAAEKGDLIAALNGIENMKDISAGDGLFNTNRH
VTNNSNMVVGONITVINAVDINNNGINISADLANVTGKNLNYLYMGYGDIALSA
NSVANNNATTEATGDLIDISKGNVNNRGNLALNGLYSVKGNLNNNDGEIRGYDV
TLATGNTDSYKSLTSETGDTTLTANVYDNAVGLAGENVSDAKSTIYNTALIAA
NKKLVINAGNLNENBDGNFLENNGALFETIDNVGIVGKEVTLTSAQVYNNNSSTI
AENGPLNLNRTGLDNTLRALLSSGADAIIRAGTGVNATVTSAGNLVYVANSNA
SDGRLEDNTATVGIASDKNLDSLVSNSYTNMISGKGVDHFNVLKGTLYNNRANLA
NALTINALNGVNFQDIAGVATLTIDTQRYTNNNSNMGLGOTTIINAVDINNRENTI
VGVSLGVKTTGNTIYNNMLSYVAGVSAKNVTNSKDAVLAGPFGALAEANEIDNT
GTIVGM"
216670..218289
/gene="ECS1283"
216670..218289
/note="probable hemolysin activator-related
protein, similar to hemolysin activator-related proteins
e.g. [Pectobacterium chrysanthemi]
gi|17726219|AAC31980.1| percent identity 27 in 484

aa, similar to hemolysin activation protein precursor
[Serratia marces]"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hemolysin activator-related protein"
/protein_id="BAB34706.1"
/db_xref="GI:13360743"
/translation="MOLRFNGLLLILCEQVAAKSIDNLIKOOKNADESTLRNKL
EKRYVNAQNRMAFTLGDLDQEKNCFTIINRELDNPLGGOLISEVOKAIGRCVGA
GVQKIAAIDQYINAGYITRISIPRODISKGLRRQVNAKTEKTIYENDISEM
LPRGRNDILNRDIEQGENLDQVAVKINIEGRDGSIVHIDHRRKNMSVRA
SYNNMGDEETGRYQFTSVGLLENPAKMGDLFYLAGTSTTGQYENVSYSIIPVGYE
YGFYSKSKSQOVPISLISLDVYIEYSAKATHVYRDKTKFAGSALIRKSN
YTLNGEELTLOARDMGNIKELANGKQOOLPGAFMTSMOERLTFWGEKTPDQOYGD
VSPSNTIFNEGENTROIENGYNVTAFAQVAPRLTLOKLTVDGRMSITGFENSVG
LSGNDGFYIKNTAFLPDKKANYALDQGOVYQDASYGDESIVGSAVGDGNIKSL
EYNFVSPTLKYPTATLIDIRVNVNFSYQW"
218305..218673
/gene="ECS1284"
218305..218673
/gene="ECS1284"
218305..218673
/note="probable holo-[acyl-carrier protein]
synthase, similar to holo-[acyl-carrier protein] synthases
e.g. [Campylobacter jejuni] gi|16968838|emb|CAB73833.1|
percent identity 39 in 121 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative holo-[acyl-carrier protein] synthase"
/protein_id="BAB34707.1"
/db_xref="GI:13360744"
/translation="MRIGTIVEIARITTKAIIINGEMAFYERYNSHRELLINPAEPY
ERASGFMAKEALVACGCGFROCDVRODIEIKHDAKAPFYVLSGRLQEIFTGGLQ
NNSLSISCHTTHAIVTYV"
218879..219448
/gene="ECS1285"
218879..219448
/gene="ECS1285"
218879..219448
/note="probable 3-oxoacyl-(acyl-carrier protein)
reductase, similar to 3-oxoacyl-(acyl-carrier protein)
reductases e.g. [Moritella marina]
gi|17221719|gb|AA042251.1| percent identity 41 in 188 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative 3-oxoacyl-(acyl-carrier protein)
reductase"
/protein_id="BAB34708.1"
/db_xref="GI:13360745"
/translation="MOCDLACQSVSALCQOIFROAGKIDVLVNNAGIVYDLSFASMS
YEFPTOYIETNMSITRLTDALMLRAARNPAIINVASIALIPSGOANYSKSKA
IITFVRLTAAEMAPWGVNAVAPGMIESKMKVKSAAVAVATSTPIRLRLKCEV
AAITFVLSASSASYIVGOTVIVDGLVMR"
219452..220000
/gene="ECS1286"
219452..220000
/note="probable (3R)-hydroxymyristol-(acyl-carrier
protein) dehydratase, similar to (3R)-hydroxymyristol-(acyl-
carrier protein) dehydratases e.g.
gi|17190847|gb|AA039621.1| percent identity 30 in 158 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative (3R)-hydroxymyristol-(acyl-carrier
protein) dehydratase"
/protein_id="BAB34709.1"
/db_xref="GI:13360746"
/translation="MSKTIIPSKIFLEMGWROPPLMVDKIADYKKGNGFVSVKVV
TYNEPYLGHFPDIPGVIIISIEIGQASEYISFLTIDIDIRREFEELSLRDIH
ARIHREMLEIIRTWRSQVGVLAQNLKFKDIAYPGDSIDVSKLAFSDASGFKHYS
VAYVGVKRLISQGTIINFRETK"

gene
CDS
EVSBDYPOASRLFCMEMLAGAPLIMDELDTGLKSLIDERSALIAQWYKSGKLAPIDPO
HLIFIMASTOYHAPVEAVNGATLRDFEVEFNQYVENQRIITIESSIRP"
/complement(1178575..182537)
/gene="ECS1260"
complement(1178575..182537)
/note="similar to PUTA_ECOLI g111787250 percent identity
99 in 1320 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="proline dehydrogenase"
/protein_id="BAB34684.1"
/db_xref="GI:13360720"
/translation="WGTTGKVKLDDATREIRKSAARIDRTPHWLIKQAFISYLEQL
ENSDFPELPALSGANESDEAPPAEPPEPDLFEADQIIPQSVSRALTAAYRRP
ETEAWSMLPEQARLPQPAEQAHKLAYQLADLRNOKNAGSRAGVQGLDEFSISSQ
EGVALMCLEALLRIPDKATRDALIRKISNGMOSHGRSPSLFVNAAWGLFTGK
LSTHNEASLSRSINRIITKSGEPLIRKGVDMARLMGEQFVGTGTIEALANRKE
EKGRYSYDMGEALTPAADQATVSYQAQTHAIGKASNGRGIYEGGISTIKLSAH
PRIHQIDRWMEELYPRKSLTTLARQYDIGINDAEADRLTSLDLKCFEPL
LAGWNGIEVIAQYQKRCPLIDYLIDATSRRLMIRLVKAYWDSIEIRADQGL
EGYPTVTRKYVTDVSYLCAKKLAVPMLIYQFATHNAHTLAAIYOLAQNYPGQY
EFQCIHGGEPELQVQKTVADGKLNRCPIYAPVGTHEHTLLAYIVRLLENGNTSP
VNRIDATSLPDELVDADPVTAVEKLAQOECOTGLPHKPTIPLPRDITGHRNSGIDL
ANERHLASLSALNSALQKNOALPMLQPVPAAGMSVNIIPAEKDIQVGRATPR
EVEQALEAVNNAPLWEPATPPAERAILHRAAVMESOMQILGILVREACKTSNAI
AEAREADFLHYAGVQDDPANEHTRPLGVQVCSIPNPFPLAFTGQIAAALGNS
VLAKPAEQTPLIAAGAILLEAGVPCPVOLGREGTGAOLGDDRVRVWMTGS
TEVATLORNIASRLDAGGRPIPIAETGMCNMIYVSSALTEBOYVIDYLAASDSAC
ORCSALRYLCODELADHTLMLKGMACEMCMNGRITTDIGCPVIDSEANVERHI
OTMSKGRPVQAVARENSEDAREMOSGFVAPTLIELDFAEILQEVGEVPLHVRN
RNQLELBOINASGYGLTGVHTRIDETIAOVGSAHVNLVYNNRVAGVAVQEP
GGEISGTGPRAGPPLVYRLANRPESALAVTLARQDAEYVPAOLCAALTOPNAL
REMANRPELOALCTOYGELOAGTORLLPPTGERNTWTLIPREYVICIADDDOAL
TOLAVALVAGSOVLPDPDALRQLVKALPSTVSERIOLAKENITTAQDFADVTHGGS
DQALALCEAVARQAGIYVSGFARGESNILLERLYIERSLSVNTAAAGNASLMTIG
"

gene
CDS
182960..184468
/gene="ECS1261"
182960..184468
/note="ECS1261"
/note="similar to PUTP_ECOLI g111787251 percent identity
99 in 501 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="major sodium/proline symporter"
/protein_id="BAB34684.1"
/db_xref="GI:13360721"
/translation="MAISTPMLVTECVYIFGMILIGFIAMSTKNPDYILGSRISGP
FYVALSAGASDMSGMLMGELGAVELSGISESTAIGLTIGAMINMKLVAGRLVHE
YNNNALTPDYFTGFEDEKSRILRIISALVILFFITTCASGIVAGARLFESTGVME
ETALMAGANAATILYFTIGFLAVSWTDVOASIMFALILTPVITISVSGFGSLTY
IKOKSINENVMKGLNFVAITISLMGCLGEGPHILARPMASSHISYHARISMT
WMITCLAGAVAVGFGIAIFREHPSVAGAVNONERVFTELAQILFNPWLAGILLSAI
LAAVMSLSCOLLVCSAITEEDLYKALRKHASQKELVWGRVAVLVAVALLAN
PENRVLIVSYAMAGFGAAGPVVLFSVMNSRMTNGALAGMITGALTIVYWKFGWL
GLYEIIGFTIGSGIYVFSHLGKAPSAQMKRPAEADAHHSAPSRLOGS"
/complement(184617..185960)
/gene="ECS1262"
complement(184617..185960)
/note="ECS1262"
/gene="ECS1262"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34685.1"
/db_xref="GI:13360722"
/translation="MRPISNLTSAATYIPPOQLSSCLETLITLVOMNNYIKRADY
STGMAVAPLVEEVHLAAMTTELRHROPOVAVSANDLOLPEPFPDVAQGITTFPK
TOEHNTNGMLKIAVSKSVCTSNITLLELHFSIVTTLMSRQVGTPTDMSVRIILD
"

gene
CDS
ESQNRVHEAAVSPAQOYGREETHFTNNGMREFCNMSPLANRVDELQOCSNTSNIP
NNAANDINNTYQTLIEDPEADPEQOYVRLTVYDMKKMAWMAQOARNALQVSGWCTRN
HENIKRANNVYLTDLHEDPEADPEQOYVRLTVYDMKKMAWMAQOARNALQVSGWCTRN
HENIKRANNVYLTDLHEDPEADPEQOYVRLTVYDMKKMAWMAQOARNALQVSGWCTRN
CKERNLAPPTARNYLKNDGLTALGELIKTGQKXWTFEFGDIYOMANNTOEERSAG
AEKSKKRGFQWSTARSLKSSQVTKQCARLAMLKNSGNSNPFYLAQPSRRT"
187142..187972
/gene="ECS1263"
187142..187972
/note="ECS1263"
/note="N-terminal part (1-35 in 276 aa) is similar to
C1016_ECOLI g111787252 percent identity 100 in 35 aa,
B1017_ECOLI g111787253 percent identity 100 in 240 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cytochrome"
/protein_id="BAB34686.1"
/db_xref="GI:13360723"
/translation="MFVPLIMLREGLEAALIVSLIASYLRKTRQGRHWIGVIGVL
AAALCLGIGIFINETGEPQKEQELFEGIVAVIAYVLTWVFMKRSKVNVOLE
QAVDSALQGRNGHGMALVMVFEFVARGLESEFPLAAFQODVGMPPGAMIGLAT
AVVGFLLWNGISRLNIGAFPKMTSLTFLVPAAGIAGAITAFHAGIGMHPQIATD
MSAVLSTSLTGCTLMGCTIMEGIFGTQEPASVSEVAVMTYILPALVAFALPPRAGATASRSA
"

gene
CDS
188030..189157
/gene="ECS1264"
188030..189157
/note="ECS1264"
/note="similar to YCDO_ECOLI g111787254 percent identity
99 in 375 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34687.1"
/db_xref="GI:13360724"
/translation="MTINFRNALQLSVAALFSSAFMANADVPQVAVTVYDKCEBM
TITVNAKGTQPTIIIONHSOKALEMELKGVWYERENIAPGFSQKMTANOPGEXDMT
CGLLTNPKRGLIYGEATADAOADALISGATTAAYKAVMAETTLQVDTKAFPTA
IKAGDIEKALVAPTRQHYERIEPIALEPESDLOGSIDAREDDYEQGAADPKFTGPR
LEKALFEDNTTKGMDQYADOLITVDYDLQKTSBELAPSPKVVGGAGALIEEVAASKI
SGEEDRYSHDLDWFOQANVEGSKIVDLRQLOKANPELAKVDANKKVVDTILAKY
RTKDFEFTYDKLTDADNNALKPITALAEIDLQLRGVIGLD"
189163..190434
/gene="ECS1265"
189163..190434
/note="ECS1265"
/note="similar to YCDB_ECOLI g111787255 percent identity
98 in 423 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34688.1"
/db_xref="GI:13360725"
/translation="MOYEDKNKGVNEPSRRRLKIGALAGSCPVAAHQKTOPAPCT
LSPVAREKQPEFGEHQAAGLITPOQAAQNSVAVPVVLAADKADLERPRLITQRRAPPT
OGGAAPETPNRPBRLDGLTIGCIYADNITLSTVSHSLFDERFGIAPMPKLOKM
TRPRLNDLALCHGDVLDIOTCANTQDTVTHALRDIIKHPDILSVMKKEGFTSDA
ARSKGKETPILNLGFKDGTANPDSQONKLMQKVWVVAQDOEPAMTIGGSYQAVRLQ
FRVEFWDRTPLEKEQTIYFGRDKQGAELGMOHEHDVDPVADSDPEGKGLADLSHILAN
PRTAESSESLMLRGYSYSLGVTNSGOLDMGILFVGYOHDLEKGFILVQKRLNGEAL
EYVAPIGGGEFFALPGVKANDYLSGSLHYV"
190778..191842
/gene="ECS1266"
190778..191842
/note="ECS1266"
/note="similar to PHOH_ECOLI g111787257 percent identity
100 in 354 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental

```
influenzae Rd) g11175795|sp|P44193|YE22_HA2IN percent
identity 40 in 158 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative anti-repressor protein"
/protein_id="BAB34674.1"
/db_xref="GI:13360711"
/translation="MTSOLIPVFNQNTIANETLLVYNARDLHTPLGVGRFASWITERI
EEGVENODYIAISOKREIGYGRKKDYHLTLDTAKETAMVERNEKGRORRYETIC
EKLIRNMQSNOAEPOQOFTDEETILLICMOJMEKAODLSKHLPIMKEILNSETTKL
YDIAPETIYMTKNDVLLREARLDQSSFVQRRAPMLKSLRAKQEEF"
/complement(172166..173560)
/gene="ECS1252"
/complement(172166..173560)
/note="similar to YCDG_ECOLI g11787241 percent identity
98 in 464 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34675.1"
/db_xref="GI:13360712"
/translation="MVRSQLRADAPCLLTSLRQLNAMEGLPHMOLKSTESGVVA
PDERLPFAQFVAVGVHNAFMGATVIMLIMGIDPNLSIMSGITLFFPITGRV
PSYLGSSAEEVAVIAATGNGOGINPISIALGCIYVYVIGIYVWKICTRMT
ERLMPVYTGAVVMAIGINLAIPAVKSSASGSPDSMAVMVYLCIGIYVTRMTOR
LLIIVGLIVACLLXGVTNVLGSKAYDFLYSHAAFMGLPHESTPAFGCAVMAELIR
VAVILVENLGLKAVAGMTGRNDPYMGRAFVGDGLATMLSGSSGVSTYVAILIG
VMAVTKVYSTLVFAAAVIAMLLGFSPFGALHTIIPAAYIGASIVYFGLIAVAGAR
IMVQNRVDLSQNGMLIMVAVTLVYGAGDFALTLGGLTIGITATFGAILNALLSR
LVDPPEVYHOEP"
/complement(173515..173973)
/gene="ECS1253"
/complement(173515..173973)
/note="similar to B1007_ECOLI g11787242 percent identity
98 in 152 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34676.1"
/db_xref="GI:13360713"
/translation="MSCGAAVNIITTDGPRAGFTASAVCSYTDPPITLVLNNG
ASVMPVFNENETLCVNTLSAGPELSNLFGKTPMEHRFAAARQNGVTCGPQLEAL
VSPFCRISQYVSGTHDILFCATIRHATPYGLVWFDRSYHALMRPAC"
/complement(174020..174610)
/gene="ECS1254"
/complement(174020..174610)
/note="similar to B1008_ECOLI g11787243 percent identity
96 in 196 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative enzyme"
/protein_id="BAB34677.1"
/db_xref="GI:13360714"
/translation="MNEAVSPGALSTLTDPARTHNGRRETPVSDETLRELYALMKWGP
TSANCSPARIVFIRTAGKERLPAISGNLOKTLTAPYATVAMDSEFERLPOLFP
HGDAKSWFTSSPOLAETAFRNSMOAAVLIICRALGLDTGPMGSGFDROHVDAPFA
GSTLKNLILNIGGDSKLPARLPRLSPFERACGLL"
/complement(174620..175420)
/gene="ECS1255"
/complement(174620..175420)
/note="similar to B1009_ECOLI g11787244 percent identity
97 in 266 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental

gene
CDS
influenzae Rd) g11175795|sp|P44193|YE22_HA2IN percent
identity 40 in 158 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative anti-repressor protein"
/protein_id="BAB34674.1"
/db_xref="GI:13360711"
/translation="MTSOLIPVFNQNTIANETLLVYNARDLHTPLGVGRFASWITERI
EEGVENODYIAISOKREIGYGRKKDYHLTLDTAKETAMVERNEKGRORRYETIC
EKLIRNMQSNOAEPOQOFTDEETILLICMOJMEKAODLSKHLPIMKEILNSETTKL
YDIAPETIYMTKNDVLLREARLDQSSFVQRRAPMLKSLRAKQEEF"
/complement(172166..173560)
/gene="ECS1252"
/complement(172166..173560)
/note="similar to YCDG_ECOLI g11787241 percent identity
98 in 464 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34675.1"
/db_xref="GI:13360712"
/translation="MVRSQLRADAPCLLTSLRQLNAMEGLPHMOLKSTESGVVA
PDERLPFAQFVAVGVHNAFMGATVIMLIMGIDPNLSIMSGITLFFPITGRV
PSYLGSSAEEVAVIAATGNGOGINPISIALGCIYVYVIGIYVWKICTRMT
ERLMPVYTGAVVMAIGINLAIPAVKSSASGSPDSMAVMVYLCIGIYVTRMTOR
LLIIVGLIVACLLXGVTNVLGSKAYDFLYSHAAFMGLPHESTPAFGCAVMAELIR
VAVILVENLGLKAVAGMTGRNDPYMGRAFVGDGLATMLSGSSGVSTYVAILIG
VMAVTKVYSTLVFAAAVIAMLLGFSPFGALHTIIPAAYIGASIVYFGLIAVAGAR
IMVQNRVDLSQNGMLIMVAVTLVYGAGDFALTLGGLTIGITATFGAILNALLSR
LVDPPEVYHOEP"
/complement(173515..173973)
/gene="ECS1253"
/complement(173515..173973)
/note="similar to B1007_ECOLI g11787242 percent identity
98 in 152 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34676.1"
/db_xref="GI:13360713"
/translation="MSCGAAVNIITTDGPRAGFTASAVCSYTDPPITLVLNNG
ASVMPVFNENETLCVNTLSAGPELSNLFGKTPMEHRFAAARQNGVTCGPQLEAL
VSPFCRISQYVSGTHDILFCATIRHATPYGLVWFDRSYHALMRPAC"
/complement(174020..174610)
/gene="ECS1254"
/complement(174020..174610)
/note="similar to B1008_ECOLI g11787243 percent identity
96 in 196 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative enzyme"
/protein_id="BAB34677.1"
/db_xref="GI:13360714"
/translation="MNEAVSPGALSTLTDPARTHNGRRETPVSDETLRELYALMKWGP
TSANCSPARIVFIRTAGKERLPAISGNLOKTLTAPYATVAMDSEFERLPOLFP
HGDAKSWFTSSPOLAETAFRNSMOAAVLIICRALGLDTGPMGSGFDROHVDAPFA
GSTLKNLILNIGGDSKLPARLPRLSPFERACGLL"
/complement(174620..175420)
/gene="ECS1255"
/complement(174620..175420)
/note="similar to B1009_ECOLI g11787244 percent identity
97 in 266 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental

gene
CDS
/product="putative acetyltransferase"
/protein_id="BAB34678.1"
/db_xref="GI:13360715"
/note="similar to B1010_ECOLI g11787245 percent identity
99 in 128 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34679.1"
/db_xref="GI:13360716"
/translation="MPKSVIIPAGSSAPLAPFVPGTLADGVVYSGTILAPDOHNVLP
ADDPKATRHVLEIRIKVIERAGSTMADYFNISFITDMKNYAAINETYAEFFPGDAP
ARFCTQGLVAPDLVETATIAHIAK"
/complement(175826..176560)
/gene="ECS1257"
/complement(175826..176560)
/note="similar to B1011_ECOLI g11787246 percent identity
97 in 244 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative synthetase"
/protein_id="BAB34680.1"
/db_xref="GI:13360717"
/translation="MPKSPVPCADSGGMMTTIARPEAITFDPOQALIVYDMONAYA
TPGGYIDLAGEVSTTRPVIANIOTATYARAAGMLIIVONGDEQYVEGAGGSPN
FHKSNALKTMRKPOLQGLKSLKSMPOYVDELVPQGDVLEPKPRYSGFDPPLS
ILRSRGIIRHIVFSIATNVCVESTLRGFFLEYGVVLEBATHQAGPEFQKALFNI
ETFRGWSVDETTCGDALSPTSFARIA"
/complement(176518..177666)
/gene="ECS1258"
/complement(176518..177666)
/note="similar to B1012_ECOLI g11787247 percent identity
99 in 382 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34681.1"
/db_xref="GI:13360718"
/translation="MONAAPRLFTFTLDBERLMMKTGVYVPGNNGULISTHAPQYMP
TFELNKAIVOKAEHYHDFALSMIKILNGFGKTEFWHNLSEFTLMAGLAVTSRIQI
YATATLTLPLPAIYARMAATIDSISGRFVYNLTQVQMKPEYEDMGIMPQGDYFSRY
DYLFTEYVLRDLVMTGKSDPKGDFEPMNOCRVSPQSPKPYICACQSDGMAFSQO
YADFNFCRGGVNTPTAPAPAPARMAQAAOTGADSVYLFVYIADETDPAARAKE
HYKAGADEBALSLTROSOKTRGSDPTNVRQANDPTSAVINMGTLVGSYASVARKL
DEVASVPGARGVLTDFDLSGETFERIOPLMQCRHLLVLTQEVAA"
177897..178535
/gene="ECS1259"
177897..178535
/note="similar to YCDG_ECOLI g11787249 percent identity
99 in 212 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tet operon regulator"
/protein_id="BAB34682.1"
/db_xref="GI:13360719"
/translation="MTQGAVKTTGKRSRAVSARKKAILSAALDTFSQGFHGRTRBOI
AELAGVSKTNLTYFPSEKALYTAIVLRQIIDIWLAPLAKAREDFAPLAAIKETIRKL
```



```

/gene="ECS1243"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1145854501gb|AAD25478.1|AF125520_73
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34666.1"
/db_xref="GI:13360703"
/translation="MFTPTFRKGTIPAIADGTIQAHDDIDEEFFQVLDGFLSKYT
PFDILHALKDVLDQRTG"
/complement(167519..167584)
/note="IS609-01, with an internal deletion"
/complement(167648..167992)
/gene="ECS1244"
/complement(167648..167992)
/gene="ECS1244"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1196325401ref|NP_049534.1| percent
identity 100 in 114 aa, similar to hypothetical proteins
e.g. yjiW protein precursor [Escherichia coli]
g1117238871sp|P52083|YjiW_EC011 percent identity"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34667.1"
/db_xref="GI:13360704"
/translation="MKRVLLIALISGVSFAGFAOOGGFQGPSEARSTVAQAEKDKDA
WVLEGSIVKKVGDREYEFKDSNLSGTIVTDIDDSIMAGQNVSPKDKVRLKEDIDLS
VEVDVKKALKLK"
/complement(168112..168324)
/gene="ECS1245"
/complement(168112..168324)
/gene="ECS1245"
/note="MokW protein (prophage maintenance; modulation of
host cell killing), identical to MokW [Bacteriophage 933W]
g1145854531gb|AAD25481.1|AF125520_76 percent identity 100
in 70 aa, also similar to Gelf [Escherichia coli]
g1117862001gb|AAC73129.1| percent"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="MokW protein"
/protein_id="BAB34668.1"
/db_xref="GI:13360705"
/translation="MLNTRCLASYVPKGRKQAMKQKAMLIALLIVICLVYATLVY
RKDLCEVRIRGTGEVAVFVDYSEK"
/complement(168558..168953)
/gene="ECS1246"
/complement(168558..168953)
/gene="ECS1246"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1145854541gb|AAD25482.1|AF125520_77
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34669.1"
/db_xref="GI:13360706"
/translation="MENEGDNITTLVQPKRDEKLNTVYTGKKNYQOSCKHRAIEV
HEODHVIILQCGCVDPFQYVLRKANDGAVVREIRQLNRHDOJRESVASLERBEK
NTKARLAAFTALLVANDLKNTEOKVNO"
/complement(169953..169612)
/gene="ECS1247"
/complement(169953..169612)
/gene="ECS1247"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1158816681dbj|BAA84359.1| percent
identity 100 in 219 aa, identical to C-terminal part of
hypothetical protein [Bacteriophage 933W]

```

```

g1145854551gb|AAD25483.1|AF125520_78 percent identity"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34670.1"
/db_xref="GI:13360707"
/translation="MLELDERERNQYIKRRDQENEBIALTVGKRLVGLGAENLLI
DSECHVALEELARDKQALLSEASERIELEAVEVSQYKAPHPQEQIADLYEMQFD
DGRICAFHTDAKAEOMQALDAGNVCQVYKLERLONLSGNSPVTGPGWISCSFRMP
DTKTAVALVAVEFDRKGRMKWATYIPHPRANGWIIIPGASMKRPSHMPLEPPQEV
N"
/complement(169633..169851)
/gene="ECS1248"
/complement(169633..169851)
/gene="ECS1248"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1158816691dbj|BAA84360.1| but
different start, similar to hypothetical protein
[Bacteriophage 933W] g1176499071dbj|BAA94185.1| percent
identity 92 in 72 aa, also similar to hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34671.1"
/db_xref="GI:13360708"
/translation="MSKIDYQALREKATKGSYIVGHTSVNQHGNLTGVCQKKK
GEPGVIAECHNQLVETDQAVATLNS"
/complement(169838..170122)
/gene="ECS1249"
/complement(169838..170122)
/gene="ECS1249"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1158816701dbj|BAA84361.1| percent
identity 100 in 94 aa, GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34672.1"
/db_xref="GI:13360709"
/translation="MNINNTTITDIALNTGIALGYPYIMFGSGRWLTLFMKMKNR
RKQQRQKAIADAEFARGIDGMEPDGPARAISRGVAILVYRSEKNEQD"
/complement(170119..170340)
/gene="ECS1250"
/complement(170119..170340)
/gene="ECS1250"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1158816711dbj|BAA84362.1| percent
identity 100 in 73 aa, similar to C4-type zinc finger
proteins (Trar family) e.g. orf39 [Pseudomonas aeruginosa
phage phi CTX] g1140638131dbj|BAA36267.1"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="C4-type zinc finger protein (Trar family)"
/protein_id="BAB34673.1"
/db_xref="GI:13360710"
/translation="MADIIDNAEIELEQNLSLQRYKSDSNAPSATHCCEGDPIDE
RRRLAVRGCRACSCOODIELINKORVK"
/complement(170388..171017)
/gene="ECS1251"
/complement(170388..171017)
/gene="ECS1251"
/note="also similar to putative phage anti-repressor
proteins e.g. [Weissaria meningitidis]
g117379691emb|CAB84545.1| percent identity 49 in 112 aa
identical to hypothetical protein [Bacteriophage VT2-Sa]
g1158816721dbj|BAA84363.1| percent identity 100 in 209
aa,
probable anti-repressor protein,
similar to hypothetical proteins H11422 [Haemophilus

```

[Bacteriophage 933W] g1|4585444|9b|AAD25472.1|AF125520_67
percent identity 100 in 133 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34660.1"
/db_xref="GI:13360697"
/translation="MAPVSGERMMNDKILMYMORVYRNSRNPENNEYKDACLKQAFRC
FEAPDGLVRSVLSADGIPYLVILGVCTGSNSVERYLPDEVKTLTFLHLAGRNAEFHT
ARRGFIKGRIGLGFERRMPDDEGFMVRIAY"
gene
156004..156660
/gene="Ecs1238"
156004..156660
/gene="Ecs1238"
/note="unknown, similar to hypothetical protein
[Bacteriophage 933W] g1|4585445|9b|AAD25473.1|AF125520_68
percent identity 100 in 218 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34661.1"
/db_xref="GI:13360698"
/translation="MGSGKSGDPTKVP7AAQIAQEEVAMKGMQDYKNILRPADNEM
EKVDILNSEQYDNIAGTTLNLGYOKOGFARKELAGNLAQSGYDPSGSRPNAMNANO
SIDVGTGIDITTRGGOVQADKYYAGIOLGALSGQKADALQSPNLSADSLAKKASD
NAAFITKQGRASLVAGLGAAGAAYAMHKRAGSGSGAGKTPGTGANAIHQHQNRL
gene
156663..157109
/gene="Ecs1239"
156663..157109
/gene="Ecs1239"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585446|9b|AAD25474.1|AF125520_69
percent identity 100 in 148 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34662.1"
/db_xref="GI:13360699"
/translation="MEYKGYETLARAGYSGADRPQGDWQTSALTRQOYDMNTRYLP
RVARLADLGENNLSIMNAQLARVGLATSSLTQAMODNOMARYGVNRPDPSNTLIG
LRNALIAGAKNGIRIEADRDQNMILLTGASAPAROKLISVGGQLVAA"
gene
157119..157370
/gene="Ecs1240"
157119..157370
/gene="Ecs1240"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585447|9b|AAD25475.1|AF125520_70
percent identity 100 in 83 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34663.1"
/db_xref="GI:13360700"
/translation="MGYGLDIANOSRRALQICSDARRRREITEANKNOMAAQOKAQ
NKONIGTIGTGAIGASVGGVAGAVATGIGIASLP"
gene
157381..158646
/gene="Ecs1241"
157381..158646
/gene="Ecs1241"
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585448|9b|AAD25476.1|AF125520_71
percent identity 100 in 421 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34664.1"
/db_xref="GI:13360701"

/translation="MSGFAQGLAGFSTVDQAMTRRKEGLREAOALAROOKNNRDE
FAOSEFHNKNVDQRFNFDYRAKVDDRNAYALKEEFENLGNQVYRNASLMEQOURLQKY
NORLEYNDMIHNSOPFLMEALGAIEGDQDEAFNGLPKGPHILMSSEGLYGA
GOAVINOKITFCDKPKMAIDSLINTPEULDIYSGFAELDORIGMPSTGDKTEAR
IGSIVPAQOEGYVILIGIDLTVSDGSTAHKVTYSGASHPDQYTLAIPVKAIAYVD
RSKFAEISKNYGYFMPKQGLSLKELQKGSNVAADIKNGGNAQAAVDEYATGSGO
PHOOKIOOKILOQOQVIMMAGDDPKLSPARNVAAQEPQKLENGYANFLRI
QKARGEQARDESASASQFIRGLKQNAQ"
gene
158716..167097
/gene="Ecs1242"
158716..167097
/note="unknown, similar to hypothetical protein
[Bacteriophage 933W] g1|4585449|9b|AAD25477.1|AF125520_72
percent identity 99 in 2793 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34665.1"
/db_xref="GI:13360702"
/translation="MAYSEORPEAOIGNONRNSLNIQGEFDSYAFSPDNRMKD
NSTFSLGDVLPFMKGFAQSVRGTMAGLGDAMIQSPVKTGARLLNFSFRRGLG
VATVODIFAGSGSGADEVIDTLPDGKAAYVDYTGKLGATGKAVSDGAKATDEMLTGK
MSGAVYALNTPMTEGYNDSAVVAKGVNLIGALVPMVAGVYARAKYDVTLRMLTA
GLEKTYIAAGMOPERATALAEAVDVKMPLFOAGLITHTSVSAQGSAMAAADAVAN
ADVSELAQSPKFOOTELATLADDPQHOLDTRQKMDLKEVADAEVDAQLTADDELLAY
NANAAKIGDQDLNLTTRGTAKTAKSVTANAEQGLTINAQGYSTQENTLRETA
GMGVSFWEGVADATIEGAAGAAGAPFGVAVADYRORROAAEETNAADATVOODDA
POPEVDPVPAQOQRESMQNNREBOLLDOYADADATGEDGSAARREASQULNELDEO
TKROAVNNELKARPSLELIEEYRRLSOKERETFEQFOALREIVYRQOEVEPKAS
OPENAEDNGSISYPTVFRPDNENVEIEINGNSRPSREIEKRPDRYPTDEKSAAG
SDYFRNAATKQKPSYVYKGENOYVAEMDPASSEDVATETITLADGERIADADPE
QPAFMDDPRRGCTGGDTVEQANLAKNAPTAELVRSQAEADGPTAQLTELTERPL
PAGDILHPGGGYPLPEGVATPDENQAGRGRTTTEGVGQSGQKQAGVAPANAGR
OGEETLRDWMVRGLIPSPDAQNATAPVREGLPADIDARNVMPQPEKILPTVRVRSLEL
AQAEVBRQAGNRDIPQETIAPESSETVSTDRATVRCGEYRKIEDFGEIEKGA
AKHRYAOLAEITLCKTLEDRDYATQPLSKPFPDVAKLAEAGADADTLAMILYRSDI
PAKTKNNTAGMGSISIKKRVHRSVEMNGYVSAKRLAEWKGMPDPRYADWOLILRPL
PSQMDASATRVYSGYQAAGKRYDPQALYSLRKNDKSNULFSESDELLAKK
VWFAEODEKSQAGDEKTAPSPDKIRFDVYNTMRSGDIEIAGKNNMRGSGFKFS
DARKYIDSHEDLVRHREKEMREISREORNAVATRGDPERRGANSPEQSDAFGR
GYQFGVNYVEPRRADLNRAYDSLHLAEVLANPTALSLNGLGLAFARGKGAAY
HYSGEVAIILTGNGDGLAHMFPSLDNYFGRYDVSNDGKLTSGDPMTEARQVYR
IFEDGRYVAEYVROEVDAFEGYIOAIKNSMPRSALLDVRSKPYVSTVEMAA
RAFERYVQDKARAGYBNDVLVIRKAPENINDNTTAIPNABELDGIRAEFHLERT
LKTRETDKGVAFYSRKGVYTRTPGNLISDYNRBAEKGSVPQVEAVARGVMSGLDQ
DLKRVVYKSOKEAELAGLELFDGYGRVHAFTYRDKREIYVADNIPDGRVREKLRHE
IHHAMEHVYTPAEYOTIITKTVLKTBDSDVATREAMRVKADAGYKESPEVQAGEFLA
HMAEKQPNKFVAMERVVALVVKVLRRGTGLKPTLEINDILVRETIITLGOVRREGYT
PREDDGASFOVRSRSGKRDPEFKYRDEGGEYRDDLAHMKSLTPTDILTVYIGTTPYL
RHIGADPLDLYISRDYVRKATNGVAKYVPMVDYIERLPELHMDPDAIYRSATERNAYM
LLDAVKNQDPPVSAVHMAKAVRSRLINKVASYGGENGKLLSMETGTLTLRREL
SRDNLHHRGLOLPRKGHSYRSADKILYLPEDIKRSGYVSTSSLSLTPEBTIASRFRQ
ODKFVYLKAVOENIRKTKGKVDSDNNAVMAEELFEHKAEMDNLAAEMERYOPLAKLLA
DYKIAQADDEYLYARHAPRNHIAKINPKMKGQSGSIGAMNSFLLMDEANAGAVENA
VEDGKGTGEMNMDLEGOQKEMDRLSLAKGKGGSIGAMNSFLLMDEANAGAVENA
RLISAYKHAARDGLSRQOASLASLKNMVTNPNRRRGGALMNSLYMFRNASISOGTANLY
RTLGHANGCEPLLERLRMKLNVPQKITALAAGAGLILSLNSVAGEDDDGYNWYDK
VPSHVERNLNLYMKSVFGKAGEYWSIPLPYGYNVFLLGHTAEGVAAADLVASRAG
NVVGVYGAIFSPIGSETSETLSGALLKNAPITLIRFPANANMENEMGAOITQENRPF
GTPKPDQSQGRSTPAYAKAFASMLNAGFSGSOYRGAVADIPEESIKFIMDITSGCTG
RFISKTITDAVYSLNGIDIPQOVPFLGIGSGVMYADQOKKYDWTETLQYHAEIK
SLTGAERTAFIDENNGKLSMNGLMQDTRRKLKDLRKORDAITADSTLSLAQSSAWYKS
VERDKIAVDRRNREYTKKVVGD"
gene
167383..167568
/gene="Ecs1243"
167383..167568

```

CDS
    /gene="ECS1230"
    149911..150153
    /gene="ECS1230"
    [note="unknown, identical to hypothetical protein
    [Bacteriophage 933W] g1|45854438|gb|AAD25466.1|AF125520_61
    but different start"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34653.1"
    /db_xref="GI:13360690"
    /translation="MLIRMSRCRAVILVOEFPEPNRIIIDSRESWLIIICDSOLGHL
    KRSMIQGRRFQNLKLEKIGVHDVALPKWEETRRQ"
    complement(150040..150282)
    /gene="ECS1231"
    complement(150040..150282)
    /gene="ECS1231"
    /note="unknown, TTG start"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative outer membrane protein"
    /protein_id="BAB34654.1"
    /db_xref="GI:13360691"
    /translation="MLPFFHCIIIAVAVATRSITPTQDVSHSHHTVHFSPPYTESYS
    LSCEFPFHMQGDIWMPFQFQIQLNKTASLIHTAH"
    150370..152073
    /gene="ECS1232"
    150370..152073
    /gene="ECS1232"
    /note="unknown, identical to hypothetical protein
    [Bacteriophage 933W] g1|45854391|gb|AAD25467.1|AF125520_62
    percent identity 100 in 567 aa, N-terminal part is similar
    to hypothetical protein [Bacteriophage P-Elbd]
    g1|75235381|gb|AF63043.1|AF151675_5 percent i"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34655.1"
    /db_xref="GI:13360692"
    /translation="MMPILAIANSIINPLIEKPAVKGISPYIDITMRCMPRVYT
    SMLEHSAVLAEDCHREFRGVITTPEROISGVKETTITPKTIFHRDPPWRAWPUVYV
    IRSPIADPHGRITVTDGRRPKVDATIAKGDGNHPTSSRLGIPAPTAQVCTVOQ
    GGSDIADPNDETRFYETETFSVDYGEGBPGPASPLETELTGTAVOLTAAPVLON
    ASIKRRRIYSASGSGADEFLVALDASVLYTDKIPAKNLGSLATMDLPPPEM
    TGLCLMANGIAAGPAGNEVSEAYLYPAMPEVNRHTTADIVAICPIGTSIVATG
    EPLYFSGVSPSTISGKIPSMOACLSRSWAMGCVLYAGTNGLSVDAANGVALAT
    EQIYSPBOMOSQFRPASIVAYPMRGETIACITKFKDGKQDYFVSPVMMDIRYLSPTED
    CAWDLAKDMRVYTGDKMSVLAGGALPSTIRWHSKIFSLPERTSFSCIRKSPAPER
    VGITIMADVPVYIHFAPGTERKGSVRLPAATGQNMVWSGFGOVERITLTSWSEMP
    V"
    152070..153338
    /gene="ECS1233"
    152070..153338
    /gene="ECS1233"
    /note="also weakly similar to tail tip fiber protein gp21
    [phage N15] g1|74446041|pir|IT13107 percent identity 24 in
    381 aa
    identical to hypothetical protein [Bacteriophage 933W]
    g1|45854401|gb|AAD25468.1|AF125520_63 percent identity 100
    in 422 aa,
    probable tail tip fiber protein"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative tail tip fiber protein"
    /protein_id="BAB34656.1"
    /db_xref="GI:13360693"
    /translation="MTRKPMRAGKDLSTVVENNEIGTGORGDRHAFVTRBELVGLK
    LARRTSGASVYALNPGEIDSTLMTVDFPKPLNFKATGSGVYLLEMDPNYRGHSL
    TEIWRGTEDDLADAVLVATTPGQYGDPPDPSGSGFTWIRFVNAAGYKKGWNAEKIQ
    AQTQIGVKAIIIDQIRDEAKSPVVSSELKREIKNAQGAVKDAAIKTTEVGLREET
    RTIGIEIRISTLDSTSESEINVDKRTTKLDEKGEAFILAMSKKAGDGTAGIGI
    VAGKDSERGPVSOVAISASOLFEPDPPNDPTAPPAVSGGKVYIPRAMYDAVETL
    VSRVVDENKAGVSISSPVRSAVINGNRQVDSQNLNIGLFSYTSQGLITRIS
    NONNGLVNRNKIEYIDNGRLAVRIGMLR"
    153353..153631
    /gene="ECS1234"
    153353..153631
    /gene="ECS1234"
    /note="unknown, identical to hypothetical protein
    [Bacteriophage 933W] g1|4585441|gb|AAD25469.1|AF125520_64
    but different start, TTG start"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative outer membrane protein"
    /protein_id="BAB34657.1"
    /db_xref="GI:13360694"
    /translation="MEYGFALYNNRVNVTGVLTPVEFLDRTAESGSKTYTNKPDCK
    SLQAVCCLEFPWNNVFADRKVKYITINDTYWMSNLEQGMGSYIYFWG"
    153637..154254
    /gene="ECS1235"
    153637..154254
    /gene="ECS1235"
    /note="unknown, identical to hypothetical protein
    [Bacteriophage 933W] g1|45854421|gb|AAD25470.1|AF125520_65
    percent identity 100 in 205 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34658.1"
    /db_xref="GI:13360695"
    /translation="MYGISTMKPDGSVWISPGFTPOCLINKGTIPATEKSPKTSIS
    GKSEPIRREKAKADWATHEQIDGYALRLHYVYKGNPEVTTYAPAFANVTPPSEY
    GIAMYNDEGENTYIGEMWLDKALIPDIFEKQDGLPCALMPALVGVIYNNKRPYDR
    PIYTTSGATGNTKITYSCSEHYSGATWDIRKPYIDKVLIVINTSYD"
    154345..155079
    /gene="ECS1236"
    154345..155079
    /gene="ECS1236"
    /note="also similar to all gene products e.g. [Yersinia
    pseudocuberculosis] g1|59027501|sp|Q569571|AIL_YERPS percent
    identity 32 in 241 aa,
    also similar to outer membrane protein low precursor
    g1|1386931|sp|P03701|VLOW_LAMB percent identity 35 in 167
    aa
    also similar to virulence protein pagC precursor
    [Salmonella typhimurium] g1|1295581|sp|P23988|PAGC_SALTY
    percent identity 29 in 180 aa
    identical to putative low precursor [Bacteriophage 933W]
    g1|45854431|gb|AAD25471.1|AF125520_66 percent identity 100
    in 244 aa,
    probable outer membrane precursor,
    similar to outer membrane protein rck [Salmonella
    typhimurium] g1|282013|pir|AA43309 percent identity 35 in
    172 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative outer membrane precursor"
    /protein_id="BAB34659.1"
    /db_xref="GI:13360696"
    /translation="MKSIATLWCAISGICAVNLSAAGGEHTISGVAHQPQGLX
    DFVKDATAHNRRETFSHFVNRYNYSLSLEYTDGRVSGTEGDKKPNQGINRYREITDD
    FGVITSTWTRSLNSQFTIDVOSADHTRKIKNPAASARTDIRANWLSLAGGSVRN
    QYSLVAMAGMGVAKVASADKIKDININSSGGSFESNSTKTSLAMAGGAQPLNESVT
    LDVAYESGSGMDRNTSCVTAGIGLKF"
    155509..155910
    /gene="ECS1237"
    155509..155910
    /gene="ECS1237"
    /note="unknown, identical to hypothetical protein
```

```
/gene="Ecs1222"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585430|gb|AAD25458.1|AF125520_53
percent identity 100 in 335 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34645.1"
/db_xref="GI:13360682"
/translation="MDPEFTGEETPEOLEKMLBGLGDVIDSDNAQDVYTEDTTEKHAD
EEAQOTQDNNVAPPTPDASVEQTQDVKEPEAKGVLRDGHVPIPEVLEAEERSGKQPA
EOEALRLRGQIAEBKRRVELLTSQIHQAGKPTPLPENKTSDEQIARIREMPELID
AVASILKRNKNYLOSRYOQSOAQOEGNGEDLSPLDAMNAVPLKTDQESDPRFSVA
VSDIGKLDNNPAMKDKTLTERFAVARARQVAVGEVSSADNKKADKTDIRKTAEEKY
KTAEOEQAVPASPDLGTTASVGTGDNFERLLGASHSEAAIMKGMNATIDALLEKL
G"
144253..145467
/gene="Ecs1223"
144253..145467
/gene="Ecs1223"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585431|gb|AAD25459.1|AF125520_54
percent identity 100 in 404 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34646.1"
/db_xref="GI:13360683"
/translation="MTTVTSAQANKLYQVALFTAAANRNSMVNLTLEQOEPAPRAVSPD
KSKTKTSACAPVVRITDLNKGQDEVTFIMHLSKRPMDGERVAREGREGEDSHAF
SLKINGRHLVDAGRMSQORTKFNLAASSARILLKTYRNDLODQCAIVHLAAGRGDV
ADDTILPTAHPPEPKTIMINDVLPPTHDRFPFGCDATSEPOIACADIFPSIGLYDNLSL
FIDEMAHPLQVRLSGDELHGEDPYVLYVTPRQMNWYTSFGKDNQMMVRAVNA
KGFNHPLFKECAMNRNIIIVRKYAGMFIIFYQGSKVLSNNLTATTKREVAANINDR
AMLLGAOLANAYGOKAGGHFNMYEKKTDMDNREIATISWINGLKKIRPEPKSGKMD
HGVIADVTAVKL"
145523..145912
/gene="Ecs1224"
145523..145912
/gene="Ecs1224"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585432|gb|AAD25460.1|AF125520_55
percent identity 100 in 129 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34647.1"
/db_xref="GI:13360684"
/translation="MAKTIIPSLSERVYTGHNESVAGCVFTVNAAEADSVIHLLS
LPVGIKRLNSQLVSTGLGTATVSIKSGEHALIDNSEAVSAKFAVYVPEPTTQRG
ELVTVTIKTAAAGTGLNLVLRITYVGY"
145962..146423
/gene="Ecs1225"
145962..146423
/gene="Ecs1225"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585433|gb|AAD25461.1|AF125520_56
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34648.1"
/db_xref="GI:13360685"
/translation="MSEKIVVYIGPKPVKKDITGSRITFPRLEPVHVSAMAMQL
GFPDVMVREHLDVLKQOQNDLROAQOAEKVLALAEAKSFVSVNGOEVDLS
KITSARLATCEAEELDIHKDPETAFAFRIRVEAARRVAETEHGTE"
146407..146970
/gene="Ecs1226"
146407..146970
/gene="Ecs1226"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585434|gb|AAD25462.1|AF125520_57
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34649.1"
/db_xref="GI:13360686"
/translation="MAKLSDFLCVVRHISGPLINMTALSMAAVAFSROSIVCRE
VTVVPAKKEITVLPBKDDECVHIITRISDNHETLVGRVYDLSGGSLSIRFACPGVY
SVLYAAVAKRGESQIPDELLTWEEVAAAGALERLPMQTGVSWSDPLRQYFVSQFSG
IRRAYRTTLATSPSSSYRNRVRORF"
146970..147620
/gene="Ecs1227"
146970..147620
/gene="Ecs1227"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585435|gb|AAD25463.1|AF125520_58
percent identity 100 in 216 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34650.1"
/db_xref="GI:13360687"
/translation="MTTITELIGRVNLTQVDPMMVRMPLOGLDYYNDVAVIILAP
DAGASLETICVVGAROVLPDGVITOLDIVCISDGSVRLSRVLDACVPEPMTKG
IPECFISNDLSPRFWLFPAPDKETSIDAVASRLPEAVYVLTQDDTPVLEERYNP
LVEMLEFRAPFSKDAAGAESGLAAQHYQSFVEQGLIKQGDALYARKKVENGGV"
147617..149554
/gene="Ecs1228"
147617..149554
/gene="Ecs1228"
[note="probable tail fiber protein, identical to putative
tail fiber protein [Bacteriophage 933W]
g1|4585436|gb|AAD25464.1|AF125520_59 percent identity 100
in 645 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail fiber protein"
/protein_id="BAB34651.1"
/db_xref="GI:13360688"
/translation="MSYVSGTILKSPDGEAISGANTITLALTYSPLASGTSASAVNR
EGGYGTMTDPEGYAVSVYKGTAVYGRIRIBESTETVILNMLRSLSVEYSIPGEL
LTFDRQIQQNVADDLATIRLNEEDTAVKNTQATQSKSAASASASDSAKTATSRAL
EAGOKATDATATATRAVTAAGNAEESSTRGSESKAGADAEKARQHAERLARLOESA
GEILKRAEATVSAEERMAENARCGPRGQETGPKGDVGPKEGTQPVQGPAGPK
GERGDVAGCAVAGPAGRCRGEKGEQEGEPQGIPLKGDTRGPRGKGQDGPKEGK
DPGCPAGPQCPKRGCRAGCQGPAGAGGEGETGPRPECPAGRCRGERGEPGPRG
EPGAGGAAVAVADATVAKQGIVOLSSATDSDDETAKATPRAVKAAMDVADAKTAE
AAAGGVGPGRGDKGDTGPPAGPAGPKDKBERGDTGVGATGEGVAGNDAGPAGPQBP
KGDGRGEGTGLGNAGPQCGPKGDTGCAAGPAGPGEGTGAAGPVCATGPGQKGP
GENTQIRFRLGPGNIIETNSHGWPFDTGALITGLTFIDPDAATVYQGFHLYVRFGD
GPNQDVKGLDVGSDPTGRTE"
149556..149825
/gene="Ecs1229"
149556..149825
/gene="Ecs1229"
[note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585437|gb|AAD25465.1|AF125520_60
but different start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34652.1"
/db_xref="GI:13360689"
/translation="MNTIKRLQMSLCCGGRHDDCENGOSITLAQLRLGPDILSDENG
IIFQDRIYQVAVLIDADKKQICQCVVAPILILRADGTWENIGMK"
149911..150153
/gene="Ecs1230"
149911..150153
```

```
/db_xref="GI:13360674"
/translation="MNMAVPPFGNSLYVHNHNEPVPYPMKPVVAGMGLAMQSLAKL
RQRPASTITTEIVMAEDGQRNMVSMPLRLAGMLOTINPKVKPEIRDKYIRYOEEC
DDVLYEWTGKGVNPRKMSVMEELNACADMKRDKNIASVFATGLNEMKQVKAHVS
KIRTVENAMNLIQFVLADTGKRTKND"
137229..137693
/gene="ECS1215"
/gene="ECS1215"
137229..137693
/note="probable endopeptidase, identical to endopeptidase
[Bacteriophage 933W] g14585424|gb|AAD25452.1|AF125520.47
percent identity 100 in 154 aa, similar to endopeptidases
e.g. Rz [Bacteriophage lambda]
g1119368|sp|P00726|ENPP_LAMB percent identity 72 in"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative endopeptidase"
/protein_id="BAB34638.1"
/db_xref="GI:13360675"
/translation="MNRVLCVYIIVLAVYGALMLATNHYRDNALTYKAQRKARL
EQAANNITDMQVROVDVADLAKYSREIADARAENETLRADVAARKRLRINATCSGT
VREATGTSGVDNATGPRLADTAERDYFILRRLITMOKLEGTOXYINEQCR"
137449..137634
/gene="ECS1216"
/gene="ECS1216"
137449..137634
/note="probable lipoprotein Rzl precursor, identical to
putative Rzl protein precursor [Bacteriophage 933W]
g14585425|gb|AAD25453.1|AF125520.48 percent identity 100
in 61 aa, similar to lipoprotein Rzl precursor
[Bacteriophage lambda] g11017781|gb|AAC48862"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative lipoprotein Rzl precursor"
/protein_id="BAB34639.1"
/db_xref="GI:13360676"
/translation="MRELKMLCVLMPLVVSACGTPPAPPCVKKPPAPAMTQPA
PDWQTPUNGIISSSERG"
complement(137725..138018)
/gene="ECS1217"
complement(137725..138018)
/gene="ECS1217"
138018..138018
/note="Bor protein precursor, identical to [Bacteriophage
933W] g14585426|gb|AAD25454.1|AF125520.49 percent
identity 100 in 97 aa, similar to Bor protein precursor
[Bacteriophage lambda] g1137520|sp|P26814|VBOR_LAMB
percent identity 96 in 97 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="Bor protein precursor"
/protein_id="BAB34640.1"
/db_xref="GI:13360677"
/translation="MKKMLLATFALALITGCAQOTFYQNKQTVARKEITTHFPVS
GIGCKTYDAKICGTENVKTEQTFVNGLIGFTLGIYTPLEARYVCSQ"
138138..138371
/gene="ECS1218"
/gene="ECS1218"
138138..138371
/note="unknown, similar to hypothetical protein
[Bacteriophage v72-Sa] g15881640|dbj|BAAB4331.1| percent
identity 85 in 75 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34641.1"
/db_xref="GI:13360678"
/translation="MNRICIPGLTMSALNGNFSAGVCGDNQKRCPTPGFLIFHEMGAI
SREACAPVMWKKPKDGTALCKYRSNMVLLCEI"
138427..139233
/gene="ECS1219"
138427..139233
/gene="ECS1219"
/note="putative small subunit terminase, identical to
putative small subunit terminase [Bacteriophage 933W]
g14585427|gb|AAD25455.1|AF125520.50 percent identity 100
in 268 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative small subunit terminase"
/protein_id="BAB34642.1"
/db_xref="GI:13360679"
/translation="MAKIDMKKLEQAFREHAETGITLDMCRKKINYNATRTIRK
GKIDHEIDKTDHEIDHIDSEPCNDGASDEKCAKSEKNCANSAETKIRGSRLL
PSNASFQSRNTHAVRREGYAKYLADNLMDDASVLEDELFTFRALSTKALKM
PADLEATDVTETRVALKDKILAKQALDRNTRARIESRTITFDYLAETPAKLRAR
ERINARDKLAEETDILTNORRGVTPVSDIVSSLHSEKNSGRLLDIPPE"
139214..140920
/gene="ECS1220"
/gene="ECS1220"
139214..140920
/note="probable terminase large subunit, identical to
putative terminase large subunit [Bacteriophage 933W]
g14585428|gb|AAD25456.1|AF125520.51 percent identity 100
in 568 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative terminase large subunit"
/protein_id="BAB34643.1"
/db_xref="GI:13360680"
/translation="MFERKNEPCDESEMEAEORFLFIMTKLSNPMWRNLNHYKION
EKGELVTFMRPAQROLFRSMHKNITILKAQOLFESFAIDYLLDOLALFPHKCGIV
AODKQASLEIFRTKIVAPDHLPMWLRSQETVYRRSGASGTYLRFHGSSIQVATSF
RSQVQILAHISEBKICAKYPAKAKELRTGTLNVSSECIIFDESTAEVGQDTEES
NRAOETIRASGLTLTAQDYKRFHYAMWQDPKYSAPVPSGLKLSKEMTYSAAVKAN
ITLIDEEKQWYINKETBQREEMKQEPSTDEAFLTGRVFSAEYTLQASFCSPM
IVYDIEPVTKATKQSLREBCKNKLRTLMVLELPPDEYVCGADTAGELH
GDRSLDVYKRSNGEYVAHFMFGHLDALFEALISQVCRNNAVGEGRNNGHVALI
KLRELPTRYLYNQHLDQVADDDPTPLGWTGTQSPVLTGCKKTLNNGISITRS
GTSLEMTYVYDAKSGMNAQBGCFDDQLMSIMIQEHRARAPRVYKQTKDKRTTHM
AH"
140920..143064
/gene="ECS1221"
140920..143064
/gene="ECS1221"
/note="probable portal protein, identical to putative
portal protein [Bacteriophage 933W]
g14585429|gb|AAD25457.1|AF125520.52 percent identity 100
in 714 aa"
/codon_start=1
/transl_table=1
/evidence-not_experimental
/product="putative portal protein"
/protein_id="BAB34644.1"
/db_xref="GI:13360681"
/translation="MKNETNYMATKNDNGCATPPRSQROLALCSDISQPKMRDANK
ACAYVGDGDLPEEVLQVLDKRGQPTIHNLIAPVVDVLGCEAKTRDLYVMSDEPD
ETEKLIANAIFPACRILGDMKNAABSDAIAEQIKALTSVEVRNRPDPGPEKYST
VSRNEVFWMLSPRADLSDCKMLMRKRMMDDEKATFPGNAOVIDIAIDMRGCVFT
TVTEGQSPSLMSAEYQSDWROQNEWLQREARRVLLQVYVYRTERLPVIELSNGV
VAFQKNMLMOVAVAASGVQVYKGRVSRIRPAAVGFHYVDRCSAPQCMFLVFPW
GYRDKDTGEPEGLISRAIIPADDEVNFRRIKLTWLOAKRVIMDDATDLSNDLMEQI
ERPDCGIKLNVPKRNKSVADVFVEDQFOVASQFQVMOSEKLIQDTMGVSAFLG
ODSGATGSVAISNLVEOGATTLAETINNYQFACQOERILLAYLLDLCKRRHNAVYI
NRDRQRRQRTVLNAEEDNELTNDISRLNTHALAAVQOTPAFKRDLAQRKSVYOG
LPFOVQAVLDDLVNLLDVPQKQEFVIRIRALSTPKSPQEMTPEEVEVAQOQALQO
QOAELOMRKQAGRAVAKLEADAARAHAQAORNSAOREVALTGOQRVYDALNOAHTAE
ITTCQNMREQEODVLAQOQMLYTLQQRNMENSL"
143222..144229
/gene="ECS1222"
143222..144229
```

gene

CDS

| | |
|--------------|---|
| gene | proteins e.g. Yjhs [Shigella dysenteriae] g1 675965 gb 1" |
| CDs | /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34630.1" /db_xref="GI:13360670" |
| misc_feature | /translation="MAFKHYDVRAAPSDLAKRITQIKLEGWQPYGALSISTFAGYGA EAFIPYVSSGSISSPEGNRPETTSAPVAPYYVIALAGNSGMYSYGEIPLDPT DSBPDIKOLARSTVTPGSAVCKYNDIIPADHGLIVQDMRLNHPKLDKSGYGT VQGLHGLAKLPEIPANAGILLVPCRGSAFTGADGTYSADASGASNSRIMVADK PLYKDIIGTKRAIKKPKNVLEFVVMKGFEPFGTPTVNAAGCALYDRKVLAD MAGCCYGSAGAGVPMICDPTTFWKQKNESTQTYGSKTKTEKNIHFPPYDENG VAVPTMKPEDDPIPEIGYYSKMRDSSATWTSQDRAHSSMARGIISDLAAVIL RHAGRALNAGASSTVEVSPSPSGSAATGYTLLSYLASESGSLKVGWASGGR AAVYSDAESTGKAVVLTKEAGKSWILEYAAAGNAGALLKQGLRCRKYGAALAN OYVMAFYVSSSLPGVALTGDGNNLLAAFYIOTDANKLNTMYNAKYNRGPFGEC FICELRIPDRFRMKHMKFLLLPEYGRAPFAIVDCITSPPTLVMPVPLHGFSSSY GVLPPLCKKH" |
| gene | complement(133196..134505) /note="I5629-02" /complement(133238..134128) /gene="ECS1208" /complement(133238..134128) /gene="ECS1208" |
| CDs | /note="transposase (insertion sequence I5629), identical to g1 17443862 pir T00240" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative transposase OrfB protein of insertion sequence I5629" /protein_id="BAB34631.1" /db_xref="GI:13360668" |
| gene | /translation="MMPLDKLREOYGVGVCSELIHAPSTYYHCOOORHHPKRSAR AHDMDIKREIORVNDENHOYGVKRVKQRLIREGIRARCVARLMAVMGLAGYLRG KRYRTTISKAYAAAGDRVNRQFVAERPDLMVADTYTSTWGEYVAFIIDVFGYI VWVRVSSSTETFDALAEQALMARPSGTIHHSKGSQYSVLATYTERKEGGLAST GSTGYSNDAEMESINGLYKAEVIHRKSWKNRAVEELATLTWDMVNNRRLIGRLGHT PRAEAKAYASIGNDILAA" |
| CDs | complement(134125..134451) /gene="ECS1209" complement(134125..134451) /gene="ECS1209" |
| gene | /note="transposase (insertion sequence I5629), identical to g1 1744868 pir T00241 percent identity 100 in 108 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative transposase OrfA protein of insertion sequence I5629" /protein_id="BAB34632.1" /db_xref="GI:13360669" |
| misc_feature | /translation="MTKNRSPSEVQRQAIRNVLSEODEYISQMAAISIAKICCTP ETLRVWRQHEHDTGGDGLTSABERQRLIKELERENRELRSNDILRQASVFAAEF DILMK" |
| gene | 134512..134778 /note="partial CDs, identical to C-terminal part of hypothetical protein [Bacteriophage 933W] g1 4585419 gb AAD25447.1 AFI25520_42 percent identity 100 in 90 aa, also similar to C-terminal parts of hypothetical proteins e.g. Yjhs [Shigella dysenteriae] g1 675965 gb AAF28123.1 AF153317 19 percent identity 53 in 83 aa, disrupted by IS insertion" |
| CDs | 134913..135092 /gene="ECS1210" 134913..135092 /gene="ECS1210" |
| misc_feature | /note="unknown, identical to hypothetical protein [Bacteriophage 933W] g1 4499806 emb CAB39305.1 percent identity 100 in 59 aa" /codon_start=1 /transl_table=11 |
| gene | /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34633.1" /db_xref="GI:13360670" |
| CDs | /translation="MTFLNQLMVFQYVVCVLLSLSGRYAMRDPFWRQIDKRAAEKFI SASOSAGSKPREPLI" |
| gene | 134959..135405 /gene="ECS1211" 134959..135405 /gene="ECS1211" |
| misc_feature | /note="unknown, identical to hypothetical protein [Bacteriophage 933W] g1 4585420 gb AAD25448.1 AFI25520_43 percent identity 100 in 148 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34634.1" /db_xref="GI:13360671" |
| gene | /translation="MCCISFRVGTGPCVTSAGDRLTKGPLKSAVSOPEANPKRSSE SGNFLNHTFRRENPMSEITSLTAEAVDVIRSEVSALOKLRHNLAEVDA ILDELGVQAPPEPPADGTTAESGEVDESPVADATEPQESVWML" |
| CDs | 135482..135697 /gene="ECS1212" 135482..135697 /gene="ECS1212" |
| gene | /note="probable holin protein, identical to protein S [Bacteriophage vW2-Sa] g1 588163 dbj BA84327.1 , similar to putative holin proteins e.g. [Shigella dysenteriae] g1 675967 gb AAF28125.1 AF153317_21 percent identity 95 in 71 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative holin protein" /protein_id="BAB34635.1" /db_xref="GI:13360672" |
| gene | /translation="MYOMEKITTGVSYTSVAGTGYWFLDLDVSPSSQMAAIGVLGSS LLFGLITFLTLNLYKRIEDRRKARAGE" |
| CDs | 135702..136235 /gene="ECS1213" 135702..136235 /gene="ECS1213" |
| gene | /note="endolysin, identical to putative endolysin [Bacteriophage 933W] g1 4585422 gb AAD25450.1 AFI25520_45 percent identity 100 in 177 aa, similar to putative endolysins e.g. [Bacteriophage H-19B] g1 4335686 gb AAD17362.1 percent identity 93 in 177 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative endolysin" /protein_id="BAB34636.1" /db_xref="GI:13360673" |
| misc_feature | /translation="MSRKLRGISAVALIAGASAPETILDOFLDEKGNHTTAYRD GAGIWTICRGATRVGDVPIPGMKLSKECKRVVAIRDALAVENKINIVPLTEPQK AGIASPCYNALGCKCPSPFYRRINNGDRKACEAIRMWIKDGRDRCIRSNMCCYGO VSRDSDSALNACWIDR" |
| gene | 136506..137075 /gene="ECS1214" 136506..137075 /gene="ECS1214" |
| CDs | /note="putative antirepressor protein, identical to putative antirepressor protein [Bacteriophage 933W] g1 4585423 gb AAD25451.1 AFI25520_46, similar to antirepressor protein Ant [Bacteriophage p22] g1 131843 isp P03037 IRANP_BPP22 percent identity 49 in 121 aa" |
| gene | /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative antirepressor protein" /protein_id="BAB34637.1" |

| | | |
|-------------|--|-------------|
| CDS | 127071..127793 /gene="Ecs1200" /note="DNA-binding protein identical to Roi [Bacteriophage VT2-Sa] g115861628 db BA04319.1 but different start,similar to Roi proteins e.g. [Enterobacteria phage HK022] g11197729 gb AC48863.1 percent identity 82 in 242 aa" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="DNA-binding protein" /protein_id="BAB34623.1" /db_xref="GI:13360660" /translation="MNELINSNAIKTMSIFIAELVSRHDKVKSITRLAVGIYRNP PMVFEKINNGLRGVEAYVEEGEGRKRSITVYAOLESPEFTARLDVRELEGATA KIPOTSEALRADLEDQKAELEKOLALAPKVEFADRVGEASGLIGNFAKVGIG PNKLEAMRDKHILIASGSRNPMQEMDRGVETVAVNTNHGIQISFTTKTGR GOOWLTRKLLDNGMLKYTREAA" 127793..128398 /gene="Ecs1201" 127793..128398 /gene="Ecs1201" /note="unknown,identical to hypothetical protein orf15 [Bacteriophage 933W] g11449798 emb CAB39297.1 percent identity 100 in 201 aa, similar to hypothetical proteins e.g. Ning protein [Bacteriophage 211] g114539482 emb CAB39991.1 percent identity 94 in 20" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="hypothetical protein Ning" /protein_id="BAB34624.1" /db_xref="GI:13360661" /translation="MAKPARCKICKEMWHPAFSNOMCCPEHGTOALERSKERE KAKRAEKKRRREOKDKLIRKLAKPRSYWIKAOQAVNAFIREDRDPCISC GTLTAQMDAGHRTTAAAPQLRFDERNIHKOCVCKQHSGLNVPYRVELNIRIGE AVDEISNHNHMTVEECRAIKAKYQOKLKLDRNSSEA" 128398..128589 /gene="Ecs1202" 128395..128589 /gene="Ecs1202" /note="unknown,identical to hypothetical protein orf16 [Bacteriophage 933W] g11449799 emb CAB39298.1 percent identity 100 in 64 aa, similar to hypothetical proteins e.g. Ning6 [Bacteriophage lambda] g119626304 ref NP_040640.1 percent identity 80 in 60 aa" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34625.1" /db_xref="GI:13360662" /translation="MTFTVKTIPDMLVEAYENOTEVARILNCSRNTVRKYTGDEKGR HAIVNGVLWYRGMGKDTDA" 128582..129016 /gene="Ecs1203" 128582..129016 /gene="Ecs1203" /note="antitermination protein Q,identical to antitermination protein Q [Bacteriophage 933W] g114585416 gb AAD25444.1 AF125520.39 but different start,similar to antitermination Q proteins e.g. [Bacteriophage H-19b] g112668768 gb AAD04655.1 percent identity" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="antitermination protein Q" /protein_id="BAB34626.1" /db_xref="GI:13360663" /translation="MKDIRQVLERMGAMANNEDYVTSPIAGFKGLIPEKYSRPQ CCDDAMVTCGCTARLRNNRDLHDLVDVYVYGFEMALARKHGCSDFICIGRLHKA EGIVEGMLMLGVRLMDRVERELPGRTSVYQKNSLRS" | gene CDS |
| gene CDS | 129265..129417 /gene="Ecs1204" 129265..129417 /gene="Ecs1204" /note="unknown,similar to part of hypothetical protein [Bacteriophage P27] g118346570 emb CAB93763.1 percent identity 89 in 37 aa,TTG start" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34627.1" /db_xref="GI:13360664" /translation="MPASPSGMLFVRSVXAMALGRRRAIGVELESGRFQTVREVN VVSNG" 129458..129533 /gene="Ecs1205" 129458..129533 /note="anticonodon: CAU, Cove score 88.46" /product="tRNA-Ile" 129458..129533 /gene="Ecs1205" 129543..129619 /gene="argN2" /note="anticonodon: UCG, Cove score 61.15" /product="tRNA-Arg" 129543..129619 /gene="argN2" 129633..129709 /gene="argO2" /note="anticonodon: UCU, Cove score 83.94" /product="tRNA-Arg" 129633..129709 /gene="argO2" 129800..130759 /gene="Ecs1205" 129800..130759 /gene="Ecs1205" /note="Shiga toxin 2 subunit A,identical to g111351074 sp P09385 SLTA_BP933" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="Shiga toxin 2 subunit A" /protein_id="BAB34628.1" /db_xref="GI:13360665" /translation="MKCIFKWLICLLIGSSVSYSEPTIDPSQOVSYSLSNIRF EISTPLEHISQGTTSVSNHTPPGSYFAVDIGLQVYQARPHRLILIDONLLYAG FVNTPATNTPRPSDFTHISVPGVTVSMTDSSYTTLQRYALERSGMOISRSLVSS YTLAMERSGNTMTKDRASRAVLRVTYAEALRRQIQREERQALSTPAVYTWTPQDV DLTLNMGRISSNVLPEYRGEDGVGRISFNINISAILGTAVAILNCHHOGARSVAANE ESQPECOITGDRVPIKINNTLMESNTAAFLNKRKSOFLYTTGK" 130771..131040 /gene="Ecs1206" 130771..131040 /gene="Ecs1206" /note="Shiga toxin 2 subunit B,identical to g111345381 sp P09386 SLTB_BP933" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="Shiga toxin 2 subunit B" /protein_id="BAB34629.1" /db_xref="GI:13360666" /translation="MKRMMAVLFALASVANAMADCAKGIIEFSKYNEDOTFVKVDG KEYWTSRMNLOPLQSAQLTGMVTIKSSPCEGSGFAEQVFND" 131527..133431 /gene="Ecs1207" 131527..133431 /gene="Ecs1207" /note="unknown,identical to N-terminal part of hypothetical protein [Bacteriophage 933W] g114585419 gb AAD25447.1 AF125520.42 percent identity 100 in 557 aa,similar to N-terminal parts of hypothetical | gene CDS |

```
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34614.1"
/db_xref="GI:13360651"
/translation="MNKQLAILEKAMDQISYALKEQALPIQTQSKIAROLCDGGF
LNEIETROMVTEKGYEINHHGIAACSHLPDDVIDEMERKQ"
gene
124008..124286
/gene="ECS1192"
CDS
124008..124286
/gene="ECS1192"
/feature="ECS1192"
/feature="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4499788|emb|CAB39287.1| percent
identity 100 in 92 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34615.1"
/db_xref="GI:13360652"
/translation="MNDYRQFENMWSKDSQFTGDELKEFAMVIMQASRAIELDI
DMPESMDFWKDEEGATAMGIEDGDKTYIAVMKAIRAGIEKNPD"
gene
124419..124634
/gene="ECS1193"
CDS
124419..124634
/gene="ECS1193"
/feature="ECS1193"
/feature="unknown, similar to hypothetical
protein [Bacteriophage 933W] g1|4499789|emb|CAB39288.1|
percent identity 97 in 71 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34616.1"
/db_xref="GI:13360653"
/translation="MSTIAELVIANFREELVRYRYRSSSLPLDELYESHSPARRYP
RDRVLRRLFKLNNEFORNRIIRSLDK"
gene
124645..124881
/gene="ECS1194"
CDS
124645..124881
/gene="ECS1194"
/feature="ECS1194"
/feature="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|5881622|dbj|BA84313.1| but
different start, also similar to hypothetical proteins
[Bacteriophage 933W] g1|4499790|emb|CAB39289.1| percent
identity 85 in 78 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34617.1"
/db_xref="GI:13360654"
/translation="MSDLSLTQPKLKECPGCGNANILMWEAGINDVWGVAECDCGA
RCAMAPSVAAAEKAMNRAGDEANLSASORSNOK"
gene
124838..125284
/gene="ECS1195"
CDS
124838..125284
/gene="ECS1195"
/feature="ECS1195"
/feature="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|5881623|dbj|BA84314.1| but
different start, also similar to hypothetical proteins e.g.
Ninb protein [Bacteriophage 21] percent identity 43 in 147
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34618.1"
/db_xref="GI:13360655"
/translation="MKQFLRLNEAIRNNAIDAILSLPIDDKSPHEVHAKPEKRSKQ
NDRMWMLNDVSRQVLHGQRLAPEDWKLDTALWIKTKLEQRSVPGIDGVVLMGV
RTSKMRKASMTLEIIMFWGSESRNVMSDSSRREYMSQRKGRAA"
gene
125281..125808
/gene="ECS1196"
CDS
125281..125808
/gene="ECS1196"
/feature="ECS1196"
/feature="probable DNA methylase, identical to hypothetical
protein [Bacteriophage VT2-Sa] g1|5881624|dbj|BA84315.1|
percent identity 100 in 175 aa, similar to hypothetical
protein Gp62 [Bacteriophage HK97]
g1|6901634|gb|AF31137.1| percent identity 98 in 175 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative DNA methylase"
/protein_id="BAB34619.1"
/db_xref="GI:13360656"
/translation="MTIKSNPAPADKQCMQPLMFDALDIERGFMLDSASDKNALC
AHMLTEADDALENSEWSHGAIWNPPISNIRPWEKAEBOCIQORQVYMLVPEDMYS
GWFSKALESDVEKRIITDGRINFIEPSTGLEKGNKSGMLLIMRPETISPRMFTVS
KAALMAIGQVRRAA"
gene
125805..125987
/gene="ECS1197"
CDS
125805..125987
/gene="ECS1197"
/feature="ECS1197"
/feature="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|5881625|dbj|BA84316.1| percent
identity 100 in 60 aa, similar to hypothetical proteins
e.g. Nine protein [Bacteriophage 21]
g1|4539480|emb|CAB39989.1| percent identity 98 in 60 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Nine protein"
/protein_id="BAB34620.1"
/db_xref="GI:13360657"
/translation="MRQRSITDILICENCKYLPTRKSRKKRPKIPESDYKTFNTYA
HLMDIRLHRRAKTR"
gene
126092..126265
/gene="ECS1198"
CDS
126092..126265
/gene="ECS1198"
/feature="ECS1198"
/feature="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|5881626|dbj|BA84317.1| percent
identity 100 in 57 aa, GNG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34621.1"
/db_xref="GI:13360658"
/translation="MFRIVCTFTSELMAVMWGRDPRRSVRLQPCVTHQTIGV
GGELRNHORASL"
gene
126262..126996
/gene="ECS1199"
CDS
126262..126996
/gene="ECS1199"
/feature="ECS1199"
/feature="probable antirepressor protein, identical to
hypothetical protein [Bacteriophage VT2-Sa]
g1|5881627|dbj|BA84318.1| percent identity 100 in 244 aa,
C-terminal part is similar to C-terminal part of
antirepressor protein Ant [Bacteriophage P22] g1|13184"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative antirepressor protein"
/protein_id="BAB34622.1"
/db_xref="GI:13360659"
/translation="MTQIIEVETLSPITNANOIPVITTELLAOLYGTPEVRIQNHEN
KVEFEGKHPFKVYNDLKEELVALNSQMLRVYILNSQNLQSLRLQISPKARSIL
LMFERGAKRAKMLFMDQADVFEKLEDYCFSCCEKTKGOEKLNLSAKENDSLVW
LMDYANRSQALFRELIPALKLIQSGYSGICHDIYGEYSYIIGRARGVLIHHTDIDILY
EPDGPTNLAWERUKNELPPSLHRY"
gene
127071..127793
/gene="ECS1200"
```


gene
CDS
/gene="ECS1183"
118452..118613
/gene="ECS1183"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34606.1"
/db_xref="GI:13360643"
/translation="MLTHMFKISMYPPIGTGHMNSIELVIGVCRHLCFLILAAALO
PLEKMYQTR"
complement(118724..119245)
/gene="ECS1184"
complement(118724..119245)
/gene="ECS1184"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g115881614|dbj|BAA84305.1| percent
identity 100 in 173 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34607.1"
/db_xref="GI:13360644"
/translation="MAKSNVQAQFKDLEELMSLNINKKATARNLKNSARLLTVQ
EEMGDVQDVLNLAERYINATEPKPSDSSITAYKSRMESAIKKFAFQSGEETPYT
PIDKSESEKDLGEPTEVECKANALHTHYDLPVLRPESGVVTITKGIPIINDITNEAE
RISILKYVVRQ"
complement(119747..120400)
/gene="ECS1185"
complement(119747..120400)
/gene="ECS1185"
/note="probable ci repressor protein, similar to ci
[Bacteriophage lambda] g1133353|sp|P03034|RPC1_LAMB
percent identity 70 in 208 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative ci repressor protein"
/protein_id="BAB34608.1"
/db_xref="GI:13360645"
/translation="WKWELARSRKKEGTQEKLAEBLQMGIGIHWLGRSRP
SDIGVFFYLGIDNISFNHDTGSPVGEYSAPVKQYEPVFSHVQAGMSPPELRTF
TKDAERLVSTTKKASDSAFMLEVGNMGTAPTGSKPSFQDMLILVDEQVAFEGDF
CIARLGDETFEKKLIRDSQVFLQPLNDQYPMICNSCSVVGKVIAGQMEETFEFG"
120518..120733
/gene="ECS1186"
120518..120733
/gene="ECS1186"
/note="probable regulatory protein, identical to
hypothetical protein [Bacteriophage VT2-Sa]
g115881616|dbj|BAA84307.1|, similar to c2 [Bacteriophage L]
g11469215|emb|CAA63999.1| percent identity 42 in 49 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative regulatory protein"
/protein_id="BAB34609.1"
/db_xref="GI:13360646"
/translation="MSNLRYRESLNISOTLLAKAVGTGGAIGHWESGRPPDLKTC
RALVACLKLGAKVSLIDVFFPEHKA"
120875..121171
/gene="ECS1187"
120875..121171
/gene="ECS1187"
/note="regulatory protein CII, identical to CII protein
[Bacteriophage VT2-Sa] g115881617|dbj|BAA84308.1| percent
identity 100 in 98 aa, similar to CII protein
[Enterobacteria phage HK022] g11631957|pir|1542398 percent
identity 96 in 98 aa"
/codon_start=1
/transl_table=11

gene
CDS
/evidence=not_experimental
/product="regulatory protein CII"
/protein_id="BAB34610.1"
/db_xref="GI:13360647"
/translation="MEOTSYSKLSQREIDRAETDILNLSTLTQRLAKMGICHSKI
SRTDWRFTASVCAFGMASDISPISRAKVALDEITKKKSPRAETDFQIDMQF"
121204..121350
/gene="ECS1188"
121204..121350
/gene="ECS1188"
/note="unknown, identical to hypothetical protein
[Enterobacteria phage HK022] g11632160|pir|1542399 percent
identity 100 in 48 aa, also similar to orf48
[Bacteriophage P22] g11871503|emb|CAA55155.1| percent
identity 85 in 48 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34611.1"
/db_xref="GI:13360648"
/translation="MTKQLSPYQDKIRHKHILDRFLSSFKQGRFRALEKVKLMQKE
KGEH"
121343..122242
/gene="ECS1189"
121343..122242
/gene="ECS1189"
/note="replication protein O, similar to O proteins e.g. O
[Enterobacteria phage HK022] g114072891|gb|AAB60272.1|
percent identity 98 in 299 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="replication protein O"
/protein_id="BAB34612.1"
/db_xref="GI:13360649"
/translation="WSNLSNLAREARLRLOQPHOSSGKVALLHKRKMDNPFYQDAE
AAHLWVHLILKAKHTPEYVMTDAGEILVGRCLLGGRLSLAETGLKDRVOYLKRF
KRLGMIDWVSHGKFSFVSVEKYDDYQSNFVADYQOITTSKAIIMPMSNTVPAYQO
ITTDKEYVNIISNTDVLSEATADKSKDPPVSCODVADVADHELPEAPKIRALNDK
RKNOJRTFMRKAGVITRDLGSGFTMQDMRNLTSLVGENCRMFEERRNHQGTVMHK
KGFDFLNDNTLYKVRGEHDDR"
122217..123668
/gene="ECS1190"
122217..123668
/gene="ECS1190"
/note="replication protein P (putative replicative DNA
helicase), similar to P proteins e.g. [Enterobacteria phage
HK022] g116863143|gb|AAf30384.1|Af069308.32 percent
identity 99 in 478 aa, also similar to replicative DNA
helicases e.g. DnaB [Escherichia coli]
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="replication protein P"
/protein_id="BAB34613.1"
/db_xref="GI:13360650"
/translation="MRVNTMTDNFYAPPSITAEQAVIGILLDDSSRQVKVLA
KPSFYSRPHKTIIFEEITRMHEQKRPVDSITFEDLEKSLTASVGCRTYIAEIKNT
PSAANIIVAMQVRETAERVAINRMTEATELLYSRNGTATOKTEAIOAITFQILDH
AKTGSRRGLRSFGEVEMWSDLEKRPDSQSGMSTGIPSLDMLSPKGVKGSFLF
VIGARPKMKKTLYLSQMALNCVAHEKRPALMPSLEPGQILEKLVGSGVNPNIYF
LPATNDADGQYOGDYDGFENRAIETANRSEITDLIYIDTTPGLSLAOTYSERRIKRE
KCGVGMILVDTYLTMTAEKADRNDAIYGTGTLKLNKLEKDCVAVILLTQNLRALES
TKRPLPSDSRPTGQIEDQCDYVGCIIHRGCAADDSPPEETELIIRLNHRGNTGYVC
IQANGAIVDTDOQSAEMRREREPEQSKKGGF"
123668..123937
/gene="ECS1191"
123668..123937
/gene="ECS1191"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g115881620|dbj|BAA84311.1| percent
identity 100 in 89 aa"

```
lambda) g1|2981722|pdb|1AVQ|A percent identity 98 in 226
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="exonuclease"
/protein_id="BAB34597.1"
/db_xref="GI:13360634"
/translation="MTPDIIILQRTGIDVRAVEQGDAMHKLRLGYITASEVHNVIAP
RSGKKWPKMKMSYFTLLAEVCTGVAPEVNAKALWAGQYENDARTLFEETSGAVIE
SPILYRDSMRKACSPQICSDNGELKCPETSPFKPRLGCFEATKASAMAVOY
SMWYTRKDAWFAVNDPDKMRBGLHYVVERDEKTMASFDEWVPEFIEKMDALAEIG
FVFGEDWR"
/complement(115351..116136)
/gene="ECS1175"
/complement(115351..116136)
/gene="ECS1175"
/note="recombination protein Bet, identical to Bet
[Bacteriophage VT2-Sa] g1|581606|dbj|BAA84297.1| percent
identity 100 in 261 aa, also similar to Bet [Bacteriophage
lambda] g1|137511|sp|P03698|VBET_LAMB percent identity 99
in 261 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="recombination protein Bet"
/protein_id="BAB34598.1"
/db_xref="GI:13360635"
/translation="MSTRALATLAKGLARVGMDSVDPEQLITTLKQTAFFKQDASQAF
IALILVAQIGLNPWTKETIAFPDKONGIVVGVDSRIINENQOFDGMDFEDNE
SCTRIYRKDNHPICVTEWMEDECRREFPKREGSEITGPMQSHKRLRHKAMTQCA
RLAFGAGIYDKDEAERIVENTAYTAERQPERDITPVNDEMQEINTLLIALDKTWD
DLPLCSQIFRRDIRASELTOAEAVKVLGFLKQKASEQKVA"
/complement(116142..116558)
/gene="ECS1176"
/complement(116142..116558)
/gene="ECS1176"
/note="host-nuclease inhibitor protein Gam, similar to Gam
proteins e.g. [Bacteriophage lambda]
g1|1381281|sp|P03702|VGAM_LAMB percent identity 97 in 138
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="host-nuclease inhibitor protein Gam"
/protein_id="BAB34599.1"
/db_xref="GI:13360636"
/translation="MDINTETETKQKSLTFPPVFLISPAFRGRYFHSYSSAMNAY
YIDRLAQSWTRHYQIAREKEAEALADMGKGLPQHLFESLCLIDHLQHGASKAI
TRAFDDVEFOERMAEHI RYVWERTIAHVOVDIDSEV"
/complement(116513..116782)
/gene="ECS1177"
/complement(116513..116782)
/gene="ECS1177"
/note="Kil protein, identical to kil [Bacteriophage VT2-Sa]
g1|581608|dbj|BAA84299.1|, similar to kil proteins e.g.
[Bacteriophage lambda] g1|138622|sp|P03758|VKIL_LAMB
percent identity 98 in 89 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="Kil protein"
/protein_id="BAB34600.1"
/db_xref="GI:13360637"
/translation="MPLOGGLLALNLKYNESPVNVYDGNALSTYLLISQESQKAD
QTLMAIOTKFTIATFIDGDKFRFAVDAYKKWILLKLRSSKSIH"
/complement(116625..116789)
/gene="ECS1178"
/complement(116625..116789)
/gene="ECS1178"
/note="regulatory protein cIII (antitermination), identical
to cIII [Bacteriophage lambda]
g1|1333661|sp|P03044|RPC3_LAMB percent identity 100 in 54
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="regulatory protein cIII"
/protein_id="BAB34601.1"
/db_xref="GI:13360638"
/translation="MQYALGWVPVACPSSESLERITRRLRDGKRLIDILNPGVPR
NCSNTGYGYP"
/complement(116862..117230)
/gene="ECS1179"
/complement(116862..117230)
/gene="ECS1179"
/note="single strand binding protein Ea10, identical to
Ea10 [Bacteriophage VT2-Sa] g1|581610|dbj|BAA84301.1|
percent identity 100 in 122 aa, similar to Ea10
[Bacteriophage lambda] g1|1376301|sp|P03757|VE10_LAMB
percent identity 99 in 122 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="single strand binding protein Ea10"
/protein_id="BAB34602.1"
/db_xref="GI:13360639"
/translation="MSNIRKYYIDYDKASIEIHDVYTEKLIHQINFMDSSEYR
LNKHSGLNAVLIIMLAQHALLIAISDDLNAVGVCFEWMNDGNGCGWPMDSSEGR
ITDIDTSGIFDSDDMTIKKA"
/complement(117413..117664)
/gene="ECS1180"
/complement(117413..117664)
/gene="ECS1180"
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|581612|dbj|BAA84303.1| percent
identity 100 in 83 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34603.1"
/db_xref="GI:13360640"
/translation="MEFHESATCDERRANANSYKPOPIAVLFTMGAMAVLCRADDTD
ARMAIGQEMENDPTNDEFITIGAPSNYLLDTCNITKKA"
/complement(117723..117995)
/gene="ECS1181"
/complement(117723..117995)
/gene="ECS1181"
/note="probable anti-termination protein N, identical to N
protein [Bacteriophage VT2-Sa] g1|581613|dbj|BAA84304.1|
but different start, similar to N proteins e.g.
[Bacteriophage 933W] g1|4585397|gb|AA025425.1|AF125520_20
percent identity 42 in 90 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative anti-termination protein N"
/protein_id="BAB34604.1"
/db_xref="GI:13360641"
/translation="MSRKTEFKGTASRRRARANLQSEATSSDKIHPTSRVYLQ
CKRKPMRAEVYITLTLTRKEGSTCLPVALIAGYRSKQLTAR"
/complement(117973..118155)
/gene="ECS1182"
/complement(117973..118155)
/gene="ECS1182"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34605.1"
/db_xref="GI:13360642"
/translation="MEOTGRLEFKORRLSTWLSQITOPHKLMDAMPQPSOEELRDC
IAKYSGGIYVQKNIT"
118452..118613
```

percent identity 100 in 95 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34588.1"
/db_xref="GI:13360625"
/translation="MATLOELIDLTPPEDEKAMNRLVKAVKDFRAAGKFFYSVLDPLLSA
YNGEIVASIDNDKGIHTASVIMPSIDAPGLTISWADWHGITLTKDGEVDDK"
complement(112172..112390)
/gene="ECS1166"
complement(112172..112390)
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585384|gb|AAD25412.1|AF125520_7
percent identity 100 in 72 aa, 5'NG start"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34589.1"
/db_xref="GI:13360626"
/translation="MPTLFREYPRKSRATEFLDILFIVLMTPI SPLIFWAIKII
ELVIELYNDVWASSENTLHNKINPYKEN"
complement(11392..112679)
/gene="ECS1167"
complement(11392..112679)
/note="ECS1167"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34590.1"
/db_xref="GI:13360627"
/translation="MVFAPKSPARSKTAGKTTCALKESSMAITASYTMLLYCQCCTD
GKYSPPGEVIGTISMACAKKEDGRIKDKTRAFAPGHKILRSNKG"
complement(112609..113076)
/gene="ECS1168"
complement(112609..113076)
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|5881600|dbj|BA84291.1| percent
identity 100 in 155 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34591.1"
/db_xref="GI:13360628"
/translation="MLRPTLIANNHRLTKSLMRTSGFGNKKVPSAPLAFASKSEVSA
CGMYRLVTVSDSPFCANCKEHHDPYKAPRYCSGTSKSEKQESKTKNRKAKIL
LVRRNAGVQWQVRLSNRMGLKYYGMDGCFCKKSAEQNRKKNHLRTKGE"
complement(112949..113722)
/gene="ECS1169"
complement(112949..113722)
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] g1|4585386|gb|AAD25414.1|AF125520_9
percent identity 100 in 257 aa, similar to hypothetical
proteins e.g. [Bacteriophage 933W]
g1|4585455|gb|AAD25483.1|AF125520_78 percent identity 95
in"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34592.1"
/db_xref="GI:13360629"
/translation="MSEINQALREKAKATKGSYIVGHTSVNQHNLGVGVCOAKK
GEPGVIAECHVNCILIESDQAVANAFIEANPATVTLALDDQERNQYIKRRDQEN

gene
CDS
percent identity 100 in 95 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34593.1"
/db_xref="GI:13360630"
/translation="MADLDSASFEELQRTATKMRRLNQAISATHCCCGDPIDE
RRRLAVQGCRCASQEDLELSKORGSK"
complement(114039..114254)
/gene="ECS1171"
complement(114039..114254)
/note="unknown, identical to hypothetical protein
[Bacteriophage 933W] but different start, also similar to
orf61 [Bacteriophage lambda] g1|508993|gb|AA96566.1|
percent identity 93 in 46 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34594.1"
/db_xref="GI:13360631"
/translation="MTPOENALRSIARQANSEIKKARQFPDPKVVNDICRSVLKHH
ETVTLGFTPTHLSTLAIGLVGFEKER"
complement(114331..114522)
/gene="ECS1172"
complement(114331..114522)
/note="ECS1172"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34595.1"
/db_xref="GI:13360632"
/translation="MHRASPYELRTSIEHSLAQIGVRFVPIPTVETDEEFHPLATSL
SQKLEMAAKAEANERDPA"
complement(114495..114677)
/gene="ECS1173"
complement(114495..114677)
/note="unknown, identical to hypothetical protein
[Bacteriophage VT2-Sa] g1|5881604|dbj|BA84295.1| but
different start, also similar to orf60a [Bacteriophage
lambda] g1|508995|gb|AA96568.1| percent identity 96 in 60
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34596.1"
/db_xref="GI:13360633"
/translation="MKHPHDNIRGAITVEYVYTKRGWVPGSLVIRNPKAORLAE
INNRKGAIVCTRHPLS"
complement(114674..115354)
/gene="ECS1174"
complement(114674..115354)
/note="ECS1174"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34597.1"
/db_xref="GI:13360634"
/translation="MSEINQALREKAKATKGSYIVGHTSVNQHNLGVGVCOAKK
GEPGVIAECHVNCILIESDQAVANAFIEANPATVTLALDDQERNQYIKRRDQEN

```
/gene="Ecs1157"  
/note="similar to C-terminal part (337-418 in 418 aa) of  
YCC_ECOLI g11787236 percent identity 89 in 80 aa, GNG  
start"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="hypothetical protein"  
/protein_id="BAB34580.1"  
/db_xref="GI:13360617"  
/translation="MKNELTALSDSICSLFTVNSNVAISLGNFSFSDLELYGYLI  
NTEGNDVYLPELPEAOAYELNKNKRNAGIQLTRAI"  
106773..108014  
/gene="Ecs1158"  
106773..108014  
/gene="Ecs1158"  
/note="similar to AGP_ECOLI g11787237 percent identity 99  
in 413 aa (Conserved in E.coli K-12)"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="periplasmic glucose-1-phosphatase"  
/protein_id="BAB34581.1"  
/db_xref="GI:13360618"  
/translation="MNKTLIAAVAGVLLASNAOQVPEGYQLQOVLMMSRHMLRA  
PLANNGSVLEOSTPNKPEWDVPGGLITKGYLEVMGYMEMLAEQGVSGECP  
PDDTVVAANSILRTVATAQFITGAPFGCDIPVHCEKMGTMDFPNVITDSAAF  
SBOAVAMEKELSKLOLTDSTYQLLEKIVNKKDSFACEKQCCSLVDGKNFSAKYOE  
EGVSGPLKGNLSLDAFTLOYTGEFPMQVAMGELKSDQOMKVLKLNKYQDLSFS  
PEVARNAVAKPLVSIIDKALVTDRTSAPKITVYLGHDSNIASLTLDAFDFKQLODNE  
RTPIGGKIYFORWDSKANRDLMKIEVYVOSAEQLRNADALTLOAPQAVTLEISGCP  
IDADGFCPMDKFPDSLNEAVK"  
complement(108052..108279)  
/gene="Ecs1159"  
complement(108052..108279)  
/note="similar to YCCJ_ECOLI g11787238 percent identity  
100 in 75 aa (Conserved in E.coli K-12)"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="hypothetical protein"  
/protein_id="BAB34582.1"  
/db_xref="GI:13360619"  
/translation="MPTQEAHNVGEMASLRMTSPRIAAIEVYAGYDEKMAKIVE  
EGSDVYLVARFATDQSLFMCQYIERKVV"  
complement(108300..108831)  
/note="partial CDS, similar to part (15-197 in 197 aa) of  
tir repressor binding protein WRBA_ECOLI g11787239  
percent identity 98 in 183 aa"  
complement(108875..110185)  
/gene="Ecs1160"  
complement(108875..110185)  
/gene="Ecs1160"  
[note="probable integrase, identical to integrase  
[Bacteriophage 933W] g1145853781gb|AD25406.1|AF125520_1  
but different start, also similar to integrases e.g.  
[Bacteriophage coli tac prophage]  
g11616234|sp|P760561|INFR_ECOLI percent identity 42 in 408  
aa"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="putative integrase"  
/protein_id="BAB34583.1"  
/db_xref="GI:13360620"  
/translation="MANSAPYAGVENHGGKLRITFKYRGKRVRENLRVPTTPKNRKA  
GELRASVCAIRGTGFDYADPPDSPNLKFLGKKDITVGLAOKWLTAKMEIGSN  
ALNRYGVKKNMPLRIGRGLASSITRKEDLPTRKDLTEKSKRSTSRKRRTYPT  
VNYITTTAGMSEFAENGITLKNPNSTIPLKSKRPDPLTRDEFSRLIDCHHOQ  
TKMLWTVAVFTGMRGEIALALAMEDIDLKAGTITVRNFKRIGDFTLPKTDACITNRYI  
HLAPALEALKNQMLRLSRHOITVOYREYGRITLIECTFVFCQIVKRNKRGIN  
YAVSSIGATWDSAIKRAIGIRSKRAYOSRHYYACWALSSCANPTFIASOMGHSASMY
```

```
gene  
CDS  
complement(110238..110522)  
/gene="Ecs1161"  
complement(110238..110522)  
/note="probable excisionase, identical to putative  
excisionase [Bacteriophage 933W]  
g1145853791gb|AD25407.1|AF125520_2"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="putative excisionase"  
/protein_id="BAB34584.1"  
/db_xref="GI:13360621"  
/translation="MRELVNOHNGIOPVTPVYVQINANEVLTLELMAVTGLRKGTI  
LRARDSAMNNGREYKOIAPDGTPEKNSCECLHPTINWIKNOPSPQDV"  
complement(110608..110919)  
/gene="Ecs1162"  
complement(110608..110919)  
/note="unknown, identical to hypothetical protein  
[Bacteriophage 933W] g1145853801gb|AD25408.1|AF125520_3"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="hypothetical protein"  
/protein_id="BAB34585.1"  
/db_xref="GI:13360622"  
/translation="MEMIMCSTFNPLTLQKYOPDPEDLCSLGGNHGKAAMECKDK  
IHLICNDVLYVDLIDREKRESEAVRALDSWRDGSAAQIYDLAISKEIPEGVRI  
E"  
complement(110979..111323)  
/gene="Ecs1163"  
complement(110979..111323)  
/note="unknown, identical to hypothetical protein  
[Bacteriophage 933W] g1145853811gb|AD25409.1|AF125520_4  
but different start"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="hypothetical protein"  
/protein_id="BAB34586.1"  
/db_xref="GI:13360623"  
/translation="MVLSKOTFGCYONRRRLTVANLQALVKGEPDAMIIRGKTEE  
YRLFNDYMKRIMFREYDRLITIKGYRKDDSSRRIDVPIYGEIKITTHPHGDKPY  
KVFARIKYNIGNE"  
complement(111256..111879)  
/gene="Ecs1164"  
complement(111256..111879)  
/gene="Ecs1164"  
[note="unknown, identical to hypothetical protein  
[Bacteriophage 933W]  
g1145853821gb|AD25410.1|AF125520_5, similar to  
hypothetical protein [Bacteriophage 933W]  
g1145854551gb|AD25483.1|AF125520_78 percent identity 50  
in 80 aa"  
/codon_start=1  
/transl_table=11  
/evidence="not_experimental"  
/product="hypothetical protein"  
/protein_id="BAB34587.1"  
/db_xref="GI:13360624"  
/translation="MTFTDKEMIKETIKERIGSLDYVDNTRRAYETALALATTEPPA  
TIDVGLIELVKGCTNFTICPDNSMSEPVPYIGLPRIDPASQPAKLSFQEWLSEORE  
KIDVDCGVSIETLTHMMKSAEAGNSPVPDWSVSCSDRMPKGGQVLLSVNFDSSL  
VEPLICARTGYSFRGDATIKPGNGIEQATHMMPLEPEPQEVNRC"  
complement(111883..112170)  
/gene="Ecs1165"  
complement(111883..112170)  
[note="unknown, identical to hypothetical protein  
[Bacteriophage 933W] g1145853831gb|AD25411.1|AF125520_6
```

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="periplasmic protein TorT"
/protein_id="BAB34572.1"
/db_xref="GI:13360609"
/translation="MRYLLFLLSLFPMAPASADNLLRMHQAHPYVAQSMPLKARY
MKLCALPSLKDSYWLISNYGMOPARARYGDIKVLVEGGYSQATQOQIDQCKQWG
AEATLLSSSTTSFPDLQKQANLPIVELNADAPVKSRYGVPQMGYQGRYLVO
WAHGRPLNVLMLPGPDNAGSKEMVEGFRRAIAGSPVAIVDIALGDNDEIQRNLDE
MLERHPEIDVAVGTAIAEAAMGEGRNLTPLTVASFHSQVRLGKRGVIMASD
QMVVQGEIAYEQAIROLOGGSVDNVSPLVLTPKNADREHIRRLSPGFRFVYFY
QHTSAKK"
/complement(98450..99142)
/gene="ECS1150"
/complement(98450..99142)
/gene="ECS1150"
/note="similar to TORR_ECOLI g111787229 percent identity
99 in 230 aa (Conserved in E.coli K-12)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="response transcriptional regulator for torA"
/protein_id="BAB34573.1"
/db_xref="GI:13360610"
/translation="MPHHIVYVEDEPYTQARLOSFTQEGYTVSVTASGAGLEIMON
OPVDLIIDINLPENGLMFLRALREKSTVGILVTSRSDIRIVGLEMGADYVRK
PLELRELVAVKNLMLRIDLARQAQPTQDNCYRFAGYCLNVSHTLERGEPIKLR
AEYEMLVAFVYNPEILISRELLMLSARVENEDLRTVDVLIIRLRHKLSADLLVQ
HGEGYFLADVC"
99272..100444
/gene="ECS1154"
99272..100444
/gene="ECS1151"
/note="similar to TORC_ECOLI g111787230 percent identity
99 in 390 aa (Conserved in E.coli K-12)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="trimethylamine N-oxide reductase cytochrome
c-type subunit"
/protein_id="BAB34574.1"
/db_xref="GI:13360611"
/translation="MRKLMALRRSPASVLAIVAGIYIGTALILVPRVGIKVTST
TERCVSHSQMPYVEEYKQSVHFRONASGVABEHCDDIPIDPEWVRKLEASDIYQ
TPIAHSIDPEKEFAKPAELEREMAKMKNNSATCSCHNYADAMHAKQHPAAROM
KVAADNOSCIDCKGIAHQLPDMSGFRKQFDELRAAGANDGDTLYSIDIKPIYAK
GDKASGSLIPASEVYIKRDGDMLOIEITGMTESAGROVLQOPEGRKIFVVASIRD
VOOQVKTEKTYVADNTNEMSKLOATPAMKKGMVNDIKPIYAVASDLYVGTGNOCHG
APEISHDANGWIGTNGMIGTISLDRKREERTLKYIQMNASDTPAGAKAGDKKEK"
100444..102990
/gene="ECS1152"
100444..102990
/gene="ECS1152"
/note="similar to TORA_ECOLI g111787231 percent identity
99 in 847 aa (Conserved in E.coli K-12)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="trimethylamine N-oxide reductase subunit"
/protein_id="BAB34575.1"
/db_xref="GI:13360612"
/translation="MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTSRRATQAQANT
EAVISKEGILITGSHWGAIRATVADGRFVAKPELILKYPKMTAGLPDHVHNAARAY
PMVRVPMLRKRLSDTSQGRDNFRVSRMDEALDMFEYEELEKQTHGPSALLTAGM
OSTGKTHNMSGMLAKAIALHGNVGTGDSYGAOAVILPRVYSGMEVYQSTSWPLV
LONSKTIVLWGSLLKNOQANMWCMPDHQDYEVYAKAKAVAAAGEIYVISTDPVYTSYH
EYLGRHVHIAVNPQTDVPIOLALAHLYSENLVKNPLANTCVGEGEOTLPYLSEK
DQPKRANAAREKTIQDAETIRGLAROMANRQIITAGWCVRMQGEGEOMAMIVVLA
AMLGQIGLPEGGFGMHWNGAGTPGRKGVILSGFSGSTISIPVHNSDVKGYSSVIP
IARFDAILLEPGKVINMNGSVKLPPLKQIFGQTHFPHHQINRIIEGMRLETVI
AIDNMTSTCRFADIVLPATTOFERNDLQYGGHNSNGIATKQAVVPDEARNDNDI
FRELCRFRNEAFTEGLDEMGLKRIWQEGVQGGRGVHLPAFDDPMNKKRYVEED

HPQMFVRHQAREDDPLEPICTPSGLIEIYKTIADNNYDDCGHPMFWFEKIERSHGG
PGSQYPLHLQSVHPDFRLHSQLESFELRQOYVAGKEPFINPDQASANGIRNGDV
VRVFNARQVLOAGVSDRYAPGARVIEGHAMHPDGKEGALCKYGNPNVLTIDIG
TSOLAQATSAHTTVEIKNGCYBQVYAFNGPVEMVAGCEYVASQVKL"
102987..103586
/gene="ECS1153"
102987..103586
/gene="ECS1153"
/note="similar to TORR_ECOLI g111787232 percent identity
96 in 199 aa (Conserved in E.coli K-12)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="part of trimethylamine N-oxide oxidoreductase"
/protein_id="BAB34576.1"
/db_xref="GI:13360613"
/translation="WTTITTAQICVYAWLAQLFSRELBQLTQIATSAQMAEFSLI
KSEPLTAADDELNRVATLVRDADARLELAADFGLMTDQKQALPVASAYKQDQO
EIKRLIVEAGNETSGNNEPDLAIYILELISHLHFSIGEGTVPARIDSLROKTLTA
LRQWLPEFVARCQYDSFGFYAALSOLLVLEEGDHQNR"
complement(103740..104045)
/gene="ECS1154"
complement(103740..104045)
/gene="ECS1154"
/note="similar to YCCD_ECOLI g111787234 percent identity
100 in 101 aa (Conserved in E.coli K-12)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34577.1"
/db_xref="GI:13360614"
/translation="MANVTYFTTTERCLHTGISEEELNEIYGLGVPEPPIQETTWY
FDDHAIVQVORAVRLRHELDMPGIAVALTIMDDIAHLQENRLLQGRSRFAHP"
complement(104045..104965)
/gene="ECS1155"
complement(104045..104965)
/gene="ECS1155"
/note="similar to CBPA_ECOLI g111787235 percent identity
99 in 306 aa (Conserved in E.coli K-12)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="curved DNA-binding protein"
/protein_id="BAB34578.1"
/db_xref="GI:13360615"
/translation="MELKQYVATMGVPTDDTKITATYRLARKYIPDYSKPEPDAAE
RREKVAAMVYLSDEQRAEYEDQMORNDPQSRQFOHGDGSEFNEDVDLFFSTLE
GQHRQSRQRPARKGHDIEIYAVFLEETLTERKRTISYMLPVYNAGMDEIETPTL
NVKIPAGVNGGQRIRLKGQTPGNGPNDGLVHIAHPPLFDIYGDLEIIVPVS
PWEALGAKVTYVTLKESILITIPGSOACQRLRVKKGGLVSKKQTDLYAVLKIWMP
PKPDENTAAALMOOLADQSSFDPKMDGKA"
105225..106124
/gene="ECS1156"
105225..106124
/gene="ECS1156"
/note="similar to N-terminal part (4-297 in 418 aa) of
YCCD_ECOLI g111787236 percent identity 85 in 295 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34579.1"
/db_xref="GI:13360616"
/translation="MGNHIGISTANNYALKQAMNDIKNEYEKKQTSITLFEVTLVCF
MRXKETRQQAADYIPCLCEBDSLEKFEETIQNDNDISLPKTLRLNDKQIYSGVGG
ITTTIQPEVDIVVLSLPGSGESITDFEGYVWGFMEYIENINACDDVYSDYCDN
DEMMSIQPEWINTPGISLDSGIDLSGISFIOSEIKKTYGLAKYAPVADGQYCLLRAIL
VLKEHESYMLGSHKTOQYVEEFKIVDKQTEALVDTAFVYLRBDVTLTFGVDSGS
DNKIQOGSLMSGNPLFFKKRIY"
106234..106482
/gene="ECS1157"
106234..106482

gene
CDS

YGESYNATGLRHVMHLRCTDTLKGPLLESYEICPIPEVVLAPEDLDVDSQRLSEVC
OMLAEPAPR"
gene
79695..81239
/gene="Ecs1134"
79695..81239
/gene="Ecs1134"
/note="similar to APPC_ECOLI g111787212 percent identity
99 in 514 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="probable third cytochrome oxidase subunit I"
/protein_id="BAB34557.1"
/db_xref="GI:13360594"
/translation="MDVIDLSRMQAFALYALHFLVPLPLGLIFELIAMETIYVVG
KTIYRDMTRFWEKFGKGINFALGVATGLMEFOGFTNSFYSNYVDIEGAPLAELAM
AFLESTFVGLFEPFGMORLNKYOHLLVTLVAFSNSFALMLNANGMOYPTGAFHD
IDLRLMETSFSELVFNPSQVRFVHVMAGVYVGMPIYMAISAWYLLRGERRVLR
SEATGVEGTALITIGTQIGDSSAYEVAQOVPYKLAHEGEMOTEPADPPHYVAME
ODQERNAPAIKIPVLLGILATHSLDKPEVPGIKNLMAETYPRLQGRMAWLLMOEISQ
NREPHVLAQAFRELEBDGYGMLLSRYAPDNMNVTAQYQAMGAIQVAVFVMSFRI
MVGCSLLLVMLIALVQLTRGKIDQHRWYKMLMSLPLFWIAIEAGMFWTEGROP
WAIDIDLPYSAHSALFTGQALFSLIMIVGLTYFLIAEYVLMQKYARLSPSAMQSD
PTQOOG"
gene
81251..82387
/gene="Ecs1135"
81251..82387
/gene="Ecs1135"
/note="similar to APPB_ECOLI g111787213 percent identity
99 in 378 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="probable third cytochrome oxidase subunit II"
/protein_id="BAB34558.1"
/db_xref="GI:13360595"
/translation="MPDYETRLFTWMLLIGVLYVEMISDGFNGICLLPLVARDDD
ERRIVNSGAWHGENOVWLLIAGLAFAMPRYAAAFSGEYVAMILLVLCSEFRPL
AFDYRGIADARWRKMDAGLVISLVPVVEYGIAGFNLLGVPAFTPLQSRVYLS
FWOLLTPPELIGLISLVNIILOGVNLQIKTVGIVHLRSQLSIKRAALLMGLCFLLA
GYMLWVIGIDGVLIAODANGSPNPLMKLVAVLPGAMNNFVESPVLNIFPLGFCPL
LTVMAIRGRNGEFLMASIMQGVITFAGITTEPPVMPSPSVSISSITLMDSTSSQL
TLSIMLVYILFPIVLYLITLMSTYKMKMGRTTETLRNEMEL"
gene
82572..83876
/gene="Ecs1136"
82572..83876
/gene="Ecs1136"
/note="similar to phosphoanhydride phosphorylase
APPA_ECOLI g111787215 percent identity 98 in 434 aa
(Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="phosphoanhydride phosphorylase"
/protein_id="BAB34559.1"
/db_xref="GI:13360596"
/translation="MKAILIPFLSLIPLPQSAFAOSEPEPELKESVIVSRHGYR
APTKATQMDVDPDAMPNMPVKGLMLTPRGELIAYIGHOROROLVADGLTKGCP
OPGOVATIIADDERKTGEAFAGLAPDCAITVHTOADSPPDLFNPDKTGSCOLD
NANTVATILSRAGSIADFTGHRQTAARELERVFNPSNCLNREKQDESCSTQAL
PSELKVSADNVSCTGAVSLASMLTEIFLLOQAQMPGEGWKRTIDSHQWMLLSLHNA
QFYLLQRTPEVARSRAIPLDLIMIALTPHPPOQAQGVTLPTSVLFIAGIDTNLANL
GGALELWMLTPGGELVEFWRMRISDNSQWIOVSLVFOQLQOMDKTPTLSL
NTPGGEVKTLTLAGCEERNAQMCSTIAGFTQIVNEARIPACSL"
gene
/gene="Ecs1137"
complement(83996..86176)
/gene="Ecs1137"
/note="similar to YCCZ_ECOLI g111787216 percent identity
99 in 726 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental

/product="hypothetical protein"
/protein_id="BAB34560.1"
/db_xref="GI:13360597"
/translation="MTTKNMTPGSTQENEIDLLRLVGLMDHRKFTIISYATFTLI
AVAYSILSTPIYQADTLVQVEKOGNALSLSLMSIPMSPESSPEIOLLOSMTIGK
TIAELNLDWVEQKYFPFVGRGAKRLTKERCELAISMNHPOLNGDOOQTLTVGEN
GHYTLBGEELFVNGVGRDLKRDGVALITADIKKPGQFVLSQRTLELAINALQETP
TVSEKSESGLELMTGDGPOLITRILNLSANNYLOQNLARQAQDSQSLFQROL
KEVSELDQAEKLNRYRQOQSDVDLNEAKAVLEQYVNDQNLNLFREAEISOLY
KDDHPTYRALLEKROTLQOEKRLNKRVSAPSQOQELRLSRDVEGRAYLOLLNR
QOELISKSSAIGNVRIIDTPAVTOPOPKPKKALNFTGLGFSVGAVALARMR
RGVAPRDLERHGSYVATIPKSEMDKRTLRKKNLFSNOQRHTKNIPIPLANPRA
DSAVEARALRTSLHFAEMETENMLMTGTATPDSGKTFVSSITAAVAQSDQVLF
DADLRGYSNMLFTVSNHGLSEYLAGDELNKYIQHKGKGFVYIRGQVPPNPESEL
LMDRLRQLELMANDHYDLVIVDPPMLVADEAAVWGRSVGTSVIRFGNLTAKESV
LSMORLEAGNIGKALIGYIKRSTAYSYGVNYYGYSYEKE"
gene
/gene="Ecs1138"
complement(86196..86654)
/gene="Ecs1138"
/note="similar to YCCY_ECOLI g111787217 percent identity
99 in 152 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative phosphatase"
/protein_id="BAB34561.1"
/db_xref="GI:13360598"
/translation="WDSMAQLKENSILVYCTGNICSPIGERILRLKRLPGYKXKSG
VHGLVKHPADATADVANNHGVSLGHAGKRLTDEMARNDLILAMSEHIAOVTATA
PEVGRKTMLEFGQMLEQKEIPDPYRRSDAFHYGVGMERASQEWAKRLSR"
complement(86630..87769)
/gene="Ecs1139"
complement(86630..87769)
/gene="Ecs1139"
/note="similar to YCCZ_ECOLI g111787218 percent identity
100 in 379 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative function in exopolysaccharide
production"
/protein_id="BAB34562.1"
/db_xref="GI:13360599"
/translation="MKRIKFEVSLTAVLSLCTACTVLPQGNLSTSNKDVIELPNQY
DLDKMNIVYPTVPTGLIDOLRAKPIMSOANPELEQOIANVEYRIGIDVAVTWVDHE
LTPPAGQYRSASDPTGNVNMNAGAIPEYIGLTKAGTTLTYOVRREITARLDSVIESO
VDVSVAAFRSQKAYVTGEVSKSGQPTITNIPITLMDAIVNAGQSLTADMDRNVLTQ
GVKTRVNLVYALMOQGDRLQRNKLHPGDLIFPRNDDLKVFYWGVEGKOSTLKMDRSGM
TLAELGNAEGMNQDVADATIGFIVIRATONQNKIANIYOLNKKDASAMITGTEFOL
EPYDIVVTTAPLARMRNVISLVPITSGVHDLTETSRIQTWPN"
gene
complement(87815..89911)
/gene="Ecs1140"
/note="similar to YMCA_ECOLI g111787219 percent identity
99 in 698 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34563.1"
/db_xref="GI:13360600"
/translation="MKRNSYLSLCLATVASSACHAEVLYTVPDPLGSSOSDGGCTGLLO
MPNARIPEGEFSVNRNDQYRYSVSLFPELGTIRITDVRTRKYSQWPEPESD
OSTYKDSFDRKLRLMEBQYMLPOYAFGRKRDLAGTGLDEGELVASKAGPFDFTLGMA
WGIAAGNAGNTNTPCIVSDKTRCHAESHDGDISFQIFRPASIPFGIEIOTFVWNL
RLKLEYGNQNDPFAKLPDASHFNVGAVYRAASMDLNSYBGRYTLMGFTLRNLN
FNDLRPALRPTPKPAYOPAPESGLQYTVVANOQLTALKYNAQFPAPETIOLRDKTLVY
GOQYKYRDSREAVRANRILVNNLPGQVEKISYQKREHMAVTTETDVASLRLQVLS
TAPQSEPLQOQREAVEDLSAFGRYSKIRIDRREFSYSPNPTLSSGSDGPEPFMFOILT
MSARVYFTDHLIDCGITFNITNNYQKRSLSLPPASTLPRTVTHTRDVTYRNDVYILN
NLQANITRADJNGNRYGOVYGGILETMTAGVSELTLYRPLDASGALVADNVIVKORDMD

```
gene
/translation="MAEHIPAPNPWPAPPPVONEOSRPLPDVAORLMOHLAENGIOP
ARNMAEHIPAPNPWPAPPPVONEOSRPLPDVAORLMOHLAENGIOPARNMAEHIPAP
PNMPAPPPVONEOSRPLPDVAORLMOHLAENGIOPARNMAEHIPAPNPAPPPVOP
NEOSRPLPDVAORLMOHLAENGIOTSKRS"
72631..72813
/gene="ECS1125"
72631..72813
/gene="ECS1125"
/translation="unknown,GTG start"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34549.1"
/db_xref="GI:13360586"
/translation="MPEEYVQPSGLVECHALODAPDAELHVEADVKKRFLXLEY
KMDVEYLIXKNFLP"
complement(73294..73476)
/gene="ECS1127"
complement(73294..73476)
/gene="ECS1127"
/translation="unknown, similar to part of avirulence protein A
(pseudomonas syringae) gii1147261spi114371AVRA_PSEEG
percent identity 468 in 56 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34550.1"
/db_xref="GI:13360387"
/translation="MSGTSGSSSDAALAVRYAEYCKWTAPGLNQAEGYKAISDLS
HHYFRAEGSSPQSFL"
complement(73581..73668)
/gene="serT"
/translation="unknown, similar to HYAA_ECOLI gii11787206 percent identity
99 in 372 aa (Conserved in E.coli K-12)"
/transl_table=11
/evidence=not_experimental
/product="hydorgenase-1 small subunit"
/protein_id="BAB34551.1"
/db_xref="GI:13360588"
/translation="MNEETFYQAMRRQGVTRRSFLKYSLATSLGLGAGMAPKIAM
ALENKPRIPIVWHLGECTCTESFIRSAHPLAKDVLISLSDYDITLMAAGTQAE
EVEFEDITIOYNGKVIILAVGNGPILGEOGMCSSGPRFEKLRKRAAGSAITAMGTC
ASMGCYQAAAPNPTQATSIDKYITTDPIIKVPCPPIDYMSAITYVMYTFDLPLVD
KMGKPLMEFGQRIHDCYRRRAHDAEGEFVOSWDDAARGCYCLYKMGCGPTTYNCS
STRMNDGSEFPIDSGHGCACENGFWMDGSEFSRYVDIPOMGTHSTADTVGLTALGV
VAAAVGVHVASAVDQRRRHHNOQPTETEHQPNEDKQA"
75209..77002
/gene="ECS1129"
75209..77002
/gene="ECS1129"
/translation="unknown, similar to HYAB_ECOLI gii11787207 percent identity
99 in 597 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hydorgenase-1 large subunit"
/protein_id="BAB34552.1"
/db_xref="GI:13360589"
/translation="MSTOYETOGVITINNAGRRLLVNDPIIRIGSHMCCEVNNIDQNYIT
NAVSGTMRGLIILGGRPDMAAFVERICGVCGYALASVYIIEAIGIKYVDN
NAIIRIMLATTMGHDHLVHFTQLAGMDNIDVLDALKAQPRTSSETLAOSLSSWPKSP
GYFFDVONRLKFEVGGQIGIFRNGYGHPOYKLPPEANLMGFAYHLEALDFOREIVK
```

```
gene
/translation="IHAFGKKNPNNWIVGMPCAINIDESGAVGAVNMERLNLVOSIITRTADFINNMI
PDALAIQOFKNPSEISGTISDSCVLSYGAEPDIDANDGERSKLMPGAVLNGFNWV
LPVDVLDPOOVQVSEIDTGRYDNDYGRPPDIDGRRPPIGDPDVPKSPDNIQOLNO
ERYISWIAPRRKGNAMVEGPLARLLIYHHGDATVSVVRMSALNLPISGISTIG
RIICRAHQAAGAKILOEFFNKLMTNLKNNLNTASTEKWPTTWPECKGVGETEAP
RGALGHMAAIIDKIDLYOCVPTTNWASPRDRGOIGAVEAALMNTKMAIPEOPLEI
LRTLSHSPDLACSTHVLGDSSELLISVQR"
77021..77728
/gene="ECS1130"
77021..77728
/gene="ECS1130"
/translation="unknown, similar to HYAC_ECOLI gii11787208 percent identity
100 in 235 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="probable Ni/Fe-hydrogenase 1 b-type cytochrome
subunit"
/protein_id="BAB34553.1"
/db_xref="GI:13360590"
/translation="MOOKSDNVVSHVFEAPVRIMHMLTVICMAVIMVTCFIGKPIP
SVSGEATLYPYKCYIRLIHRSAGNVTYVILMTIYAFVGNKRSRELFYIPVRRKSW
QGVWYELRWYLFIAKPSADIGNPILQAAMFQYFLMSVFMITTGELVSEHSQYALF
APERYVEFFYWTGNSMDIHSWHRGLMGLIGAFVIGHYMALREDIMSDDTISTWY
NGYRSHKFGKISNKR"
77725..78312
/gene="ECS1131"
77725..78312
/gene="ECS1131"
/translation="unknown, similar to HYAD_ECOLI gii11787209 percent identity
98 in 195 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="processing of Hyaa and Hyab proteins"
/protein_id="BAB34554.1"
/db_xref="GI:13360591"
/translation="MSEORVYVGLNLMADEGFGVRAERLYAHYHMPSEYIYDG
GTGGLNLGIVFESASHLITDAIDGLFPGTATYGERIPATYSKKMSLHONSSE
VIALADIRGLPHAHNLVNGQPMADDYGSSELSARLEQPLPAEQALADLAAGTVP
OPANESRCLNYDCLSMENTEGVRLROYRMTEBOG"
78309..78707
/gene="ECS1132"
78309..78707
/gene="ECS1132"
/translation="unknown, similar to HYAE_ECOLI gii11787210 percent identity
99 in 132 aa (Conserved in E.coli K-12)"
/transl_table=11
/evidence=not_experimental
/product="hydorgenase 1 formation factor Hyae"
/protein_id="BAB34555.1"
/db_xref="GI:13360592"
/translation="MSNDTPEFALWQRMARGWTPVSEFRLDWLTQAPDGVLLSSD
PKRTPEVSDNPWVIGELIRFEPDYQVVAIDLEOSEAIGDRGVFPATVLTFTGNG
YKGVGLGHPMALMLIMMGLIVEPOOERAS"
78704..79561
/gene="ECS1133"
78704..79561
/gene="ECS1133"
/translation="unknown, similar to HYAF_ECOLI gii11787211 percent identity
99 in 285 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hydorgenase-1 cofactor biosynthesis protein
hyaf"
/protein_id="BAB34556.1"
/db_xref="GI:13360593"
/translation="MSETFPHLLGRTPOPNDSFSNPLPITGVNDEPSMAALQQA
HSPOVALNLNEOHOSQOPPLGVEVLAVDLNLNADNRHFTNLLGSESVYRQQA
DDSESEIQAIFCGLMVRVRRGEKLEDEKLEAGCAPLALWQAATQNLPLPDSLPPP
IDGLMNGPLLAHELLAHVNPDAOPHSINLTOLPISADRLFLSRICGGINQIQTIG
```


/product="putative tail assembly protein"
/protein_id="BAB34541.1"
/db_xref="GI:13360578"
/translation="MATTNAPSLASPPRLICLHGDLQREGRRSLVYNTAAEALRAL
SMQMPGERKQNEGWYQIRIAGDTPAPVAYARLHEQEGEVIIHIVRLAGAKGGL
QIVLGAALVSGFTASGALMGALGAFSATMTLFSGASMIIGVQOMLAPKA
KTPDRAIDNGRONVTFSSLDNMIAGNPMVPYVGYEMLVGSRISODISTDEGGDC
VVVIGRQA"
gene
65274..65480
/gene="Ecs1119"
65274..65480
/gene="Ecs1119"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34542.1"
/db_xref="GI:13360579"
/translation="MKAVTGRWMLSGGRHKSEKIPQCSRTGTAGALRLSVRNTSYVT
TKTLTGRCMCRSROGERIYRPEE"
CDS
complement(65511..66038)
/gene="Ecs1120"
complement(65511..66038)
/note="copper/zinc superoxide dismutase, similar to
copper/zinc superoxide dismutases e.9. (Salmonella
typhimurium) g1|246269|emb|CAA7358.1| percent identity
58 in 175 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative copper/zinc superoxide dismutase"
/protein_id="BAB34543.1"
/db_xref="GI:13360580"
/translation="MKCKIIAIAIMLPAACSGVAEOEVPNNVLSADGKEVSTKINT
QETPYGLPDLPHSLSEGIHGFIHVEKGCAPALKDGPVALISAGHDPKWTGCH
LGPMSPGHPGDLPALEVTHTDKANVYLAURLNSLKEIKGRSLMLAGDNDHDE
PLGGGARACGIIQ"
gene
66172..69645
/gene="Ecs1121"
66172..69645
/gene="Ecs1121"
/note="probable host specificity protein, similar to host
specificity proteins e.9. Gpu (Bacteriophage lambda)
g1|138412|sp|P03749|VHSD_LAMB percent identity 65 in 1156
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative host specificity protein"
/protein_id="BAB34544.1"
/db_xref="GI:13360581"
/translation="MGKGGRAHPVEAKDNLSTOMMSVIDAIGEGPIEGPVKGIOS
ILVNTPLPTDGNPVHGVTAVRAGSEORPPEGESSGAFALGEVETKAKPYTR
TITSANIDRLRYFGVQSLETTSKDRNSTRLLIOLDRNGWTEKVTNGKTT
SOFSLAYLDNLPREPPIRMVRETDASTDOLQNTLMSYEIIIDVQCYNATAY
GLQVDAEQEGGQOMVYHIRGIIQVPSNYDEPKRTSGIMDSLKPAYSNPAMCL
MDMLTPRYCGKRLGADVDKMLALYAIOYCOMVDEGGEGEPRTFAVYSOOR
AMVLSDECSAMRCMPYMNCGOTLFPVDSDFVDDVVDNGVDFVRSFSL
KDRHTAVEVNYTPONGMORSTELVDEPAIIYGRMLLMDAFCTSRQARAGIM
VITGILLETOTVPTLGSQGLRTPDITICNDRIAGTGTGRVLSDASSTLTD
REVTLPETGATVNLINGSKPVSAITAHAPDRIGVSTLPDGVETGVAGSLSTLD
RRRLFCVSTIRENTDFTAITAVOHVEKRALVNGARFEPQSTLNSVTPAVOHLT
VEVSADGVOILAOKMDTPRVVGVRSRLITSGSGESRLVTTATLADTEHSSGIL
PGEYTLTFRALINSYGOGEPAATTPRINAPVAPATIELTPGFOITAVPRLAYDPLV
OPEFMSSEKTIADTSOETSARVLTGSGOMSVSGPHIKKQKDPFVYRSNVLKGSF
VEASGRASNDABEYGLIFREKIGKHLAOLMELINSQLADMAEKRSITETRNEI
TOTVSKTLEDOKAIIOQIDRVQKTDNDIALMLVYQTKNGIPYVAGIGAEIDTD
GQPLSNILLLADRIAMIPEDGWTFPLFAQGNQLFMDVFLKRLFAVSITSSANPT
FSLTPGGRILARNADISGNVANSGLTNNTVINENCRLVGLKLSANOIEGLVLTVGKA
FPDSDRAPEWRMPSGTLTVRYDDQPDFRQIVIPAVAFSGAKHREHDTDISSRLIVR
KNGAEIYNRTALDNTLIYSGVIDMPAGHGMTEFSVSAMLVNNMPTPASISDLVYV

gene
69713..70312
/gene="Ecs1122"
69713..70312
/gene="Ecs1122"
/note="probable outer membrane protein, similar to Lom
outer membrane proteins e.9. (p9phage P-ElbA)
g1|7532789|gb|AA663231.1|AF151091_2 percent identity 68 in
199 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative outer membrane protein"
/protein_id="BAB34545.1"
/db_xref="GI:13360582"
/translation="MKRLYAAILSAICLAAGAPAMASEHOSTLSAGYLVHSTNVP
SDELNGINVKRYREFTDGLGWTSESYAGDKNRQLTFTSDTRHEDSVRRMWFVVMG
PSVRVNEFSAVAMAGVAYSRSVTFSGDYLRYTPNKKRTHDVLTGSDGGRHSNTSLAM
GAGYQFNPDESVAIDIAIEGSGSGDMRTDGFIVGVYK"
CDS
70371..71690
/gene="Ecs1123"
70371..71690
/note="probable tail fiber protein, similar to putative
tail fiber proteins e.9. [bacteriophage 933W]
g1|4585436|gb|AMD25464.1|AF125520_59 percent identity 38
in 370 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative tail fiber protein"
/protein_id="BAB34546.1"
/db_xref="GI:13360583"
/translation="MNAVAKISGVLDKDGTPVENCITOLKARRNSANVYVNTVASN
PDEAGRYSMDEVGOYVILLVEGFPSPHACTIVYDSOPGTINDPLGAMSEDDVPR
EALRPFILWBEARHAEBAKKNKGEAETSRRNAGISASQESASANAADVSAGDSAV
ARQAESAAAKOSSEASSSASAAKSESSQSAEALSRKTAASAGNARADT
TATEKAREASAEASQSRJAABEAVNRKIPTVVGPGPKPEPAPQPGKRGKGR
GDTGPAGATGERGAGDAGPAGPQGGDGERGETLTGNAGCPQKPGDGAAGPAG
POGPKGTGAAGVATGATGPOGPKDPEPTOIRFLPLGASIIETSHGMPFOTDCAIT
GLPFLAKDNTARVQFFQHLQVRRGDDPMDQVGLDVGSDPTGRTE"
gene
71692..71961
/gene="Ecs1124"
71692..71961
/gene="Ecs1124"
/note="unknown, similar to hypothetical protein
[bacteriophage 933W] g1|4585437|gb|AMD25465.1|AF125520_60
percent identity 938 in 129 aa, also similar to the
C-terminal part of putative tail protein of 933W
g1|4585436|gb|AMD25464.1|AF125520_59 percent identl"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34547.1"
/db_xref="GI:13360584"
/translation="MNLKIKIMORLGGCGKHDCHEGQSLLVQLRLGPADILSDENG
ITPEODGVTQVYVILADKKQICQVYRPLQIILADGKWMENIGMK"
CDS
complement(72086..72655)
/gene="Ecs1126"
complement(72086..72655)
/note="probable secreted effector protein,
similar to EspF proteins e.9. [Escherichia coli strain
E2348/69] g1|2865308|gb|AAC33840.1| percent identity 37 in
87 aa, also similar to L0016 - Escherichia coli
g1|3414884|gb|AAC31495.1| percent identity 38 in 126 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="EspF-like protein"
/protein_id="BAB34548.1"
/db_xref="GI:13360585"

```
gene
58914..59519
/ gene="ECS1111"
/ note="tail length determinant, similar to tail length
tape measure proteins e.g. GpH [Bacteriophage lambda]
g1138843|sp|P03736|VMTG_LAMB percent identity 77 in 859
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative tail length tape measure protein"
/protein_id="BAB34537.1"
/db_xref="GI:13360574"
/translation="MSQPVGDVLTIDSLDAVFRDEQMSVRRHFSGLTDVVKTSASV
EGLSRQALAAQAGISVQYKAMRTLPQFTDIATQLAGONPWLILOOGGVQVDS
SEGMIPEMRGLAGAITLPMGVSTSLAVATGALVAMYGDSLTLSFNKTVLSGKOS
GLTADRMILTISRAGQAGLTENQRESIALVNAVGSGEOPDAINOSVARRASAGV
EVDKVAEANGKLTLPDPTSGLTAMVQFRVNTAEQIAYVQLORSQDEACALANDIA
TKGFDEQTRRLEKNGTLETADTKGAKKSMMDATLIDIGRESSADMILASKOKARDE
ADKKMOWYQSRSQRGKTASFRANLQGANWDRNARLGLAAATLQSDMEKAGELARD
RAERDASQIKYTGAEQAKAYERLUTPLEKTYARQEEINKALDKGKILRADYNLTMAAK
TGKTPVAAYVSPVYEGSKVLRRHGGETRVLKPGYKPKHEFPMSR"

gene
58914..59519
/ gene="ECS1111"
/ note="unknown"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="hypothetical protein"
/protein_id="BAB34534.1"
/db_xref="GI:13360571"
/translation="MTAESYDDVYLDDEDADWTATQGGKSAQDTSFTLAMPKEBQ
KGLIGWESGDYRAKIRFPNGTVDFRGWSSIGKAVTAKEVITRTVATVNGKPSV
AEERSKITPVSAIKVPTSGTAKGKTTLTVSFEESATDKTEFRAVSDPSKATISV
KDMITVNGAVGKQVQIPVSGNGQFAVAAYVTTEAGAG"
59461..59964
/ gene="ECS1112"
/ note="probable tail component, similar to minor tail
proteins e.g. GpG [Bacteriophage lambda]
g1138842|sp|P03734|VMTG_LAMB percent identity 68 in 143
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative minor tail protein"
/protein_id="BAB34535.1"
/db_xref="GI:13360572"
/translation="WVSSPWIKSPLLKRALQGRKRSYFLKTEQEVNGSVTLSEL
SALQRIEHLALKRAEQAESGNLQVSEDIYRTGAFVAMSLMHNHPOKQTPSMN
EAVMKTEQEVLTWPADATARAEDVVLCLSGSGAVRPDITIEVAKNNMTLDDPSA
GKSSTAS"
60015..60404
/ gene="ECS1113"
/ note="probable minor tail component, similar to minor tail
proteins GpC-T [Bacteriophage lambda]
g11742917|pir|ITLBPTL percent identity 72 in 124
aa, probably produced by translational frameshift"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative minor tail protein"
/protein_id="BAB34536.1"
/db_xref="GI:13360573"
/translation="MLAGMTSTFYADMRHFRYRTYFPQDTQLDMHFSGLTYAVLSLFFC
DDMHPSDFSLVPRHEEQVERPDEDKMLQKAKAGLAGVFRGSDGRDILSSADVA
DYVWDPAALMMASAGIPGCVRYVPAGW"
60385..62964
/ gene="ECS1114"
/ note="tail length determinant, similar to tail length
tape measure proteins e.g. GpH [Bacteriophage lambda]
g1138843|sp|P03736|VMTG_LAMB percent identity 77 in 859
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative tail length tape measure protein"
/protein_id="BAB34537.1"
/db_xref="GI:13360574"
/translation="MSQPVGDVLTIDSLDAVFRDEQMSVRRHFSGLTDVVKTSASV
EGLSRQALAAQAGISVQYKAMRTLPQFTDIATQLAGONPWLILOOGGVQVDS
SEGMIPEMRGLAGAITLPMGVSTSLAVATGALVAMYGDSLTLSFNKTVLSGKOS
GLTADRMILTISRAGQAGLTENQRESIALVNAVGSGEOPDAINOSVARRASAGV
EVDKVAEANGKLTLPDPTSGLTAMVQFRVNTAEQIAYVQLORSQDEACALANDIA
TKGFDEQTRRLEKNGTLETADTKGAKKSMMDATLIDIGRESSADMILASKOKARDE
ADKKMOWYQSRSQRGKTASFRANLQGANWDRNARLGLAAATLQSDMEKAGELARD
RAERDASQIKYTGAEQAKAYERLUTPLEKTYARQEEINKALDKGKILRADYNLTMAAK

gene
62961..63290
/ gene="ECS1115"
/ note="minor tail component, similar to minor tail proteins
e.g. GpM [Bacteriophage lambda]
g1138845|sp|P03737|VMTM_LAMB percent identity 82 in 109
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative minor tail protein"
/protein_id="BAB34538.1"
/db_xref="GI:13360575"
/translation="WKTFRKVKPDEVNSQPSVREYRGDGYSQRMAGLNDLKY
RVTLSVTRERARHLEAFLEHGGWKAFLWTPYAMRQIKVTCAWSSRVRLRVERSA
EFKQVYN"
63290..63988
/ gene="ECS1116"
/ note="minor tail component, similar to minor tail proteins
e.g. GpL [Bacteriophage lambda]
g1138844|sp|P03738|VMTL_LAMB percent identity 76 in 232
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative minor tail protein"
/protein_id="BAB34539.1"
/db_xref="GI:13360576"
/translation="MODIHESINSESVKSPQSRVYLREITLVOGGERYFCNEINE
KGAIVTWQGHQVYRPIDQSGFPMNGKSSARSLVYSNLFGLVTAQAELOSLVAT
VWRRVYARFLDYNFVAGNEADPQELSDRWVVEQMSLTMVTAFLVATPETYD
ALEPGRIMLANTCMTYRSDCEGYTGAVADEPDKPTTDIRKDRCSKCMRGKRWY
ANFGFLSINKLSQ"
64107..64742
/ gene="ECS1117"
/ note="tail assembly protein, similar to tail assembly
proteins e.g. GpK [Bacteriophage lambda]
g1139638|sp|P03729|VTK_LAMB percent identity 84 in 196
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative tail assembly protein"
/protein_id="BAB34540.1"
/db_xref="GI:13360577"
/translation="WNISAPPEAYFRITAEPMWIRAEQGTVALVYSHRQGLFWLSEA
DRRLQIKSALSWMLVCGEITHFRVPHLGLRLEHGVTDCTYLPFDAYHLAIDMPD
FEREDDMWRNGONLVLDMNEATGFRISLPSAQPDQILCCFASVANAHAIVCGNCE
LHHLPEQLSKRRSEKQWRTHSAMRHRHWHVSAFTGIYNDLAASACM"
64640..65320
/ gene="ECS1118"
/ note="tail assembly protein, similar to tail assembly
proteins e.g. GpI [Bacteriophage lambda]
g1139637|sp|P03730|VMTI_LAMB percent identity 68 in 224
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
```

```
IFRGSSGQVALLNSTGDGLCLTLKRLERGRFAMPSPARDGKVFITQAOALMLEGIDW
RQPKRLTSLTML"
gene
51085..52623
/gene="Ecs1104"
51085..52623
CDS
/gene="Ecs1104"
/translation="unknown, similar to hypothetical proteins e.g. L0015
[Escherichia coli O157:H7 strain EDL933]
g1j341483|gb|FAC31494.1| percent identity 100 in 512 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34527.1"
/db_xref="GI:13360564"
/translation="MNDISDDIFILKORLAEOBALIHAIQEKLSNREIDHIOAOL
DKLRMNGSRSEKVRRIAQMEDLNLOKESDPTLGRVYDPAVORPLROTFRKRP
PESLPRDEKLLPAPCCPCNGGSLSTIGETPAOLEMBSAFYTRIVERKHACTOC
DAIVAPAPSPRTERRIAGPGLARVLTSKTAETPLYRQSEITYRGQVLEKRSLSG
WVDAACRLSPLEBLHGYVMTDGLADDPVQVLLPGNKRTGLRMAYRDRNA
GSLAPAWFAVSPDRKGIHPQTHACFSGLQADAYGPNELYNRGITPAACWAHA
RRKIHDAVRIPSALETBALQIGOLYAIADIGMPAEORLAEROKTRKLSLES
WLREKMTLSRHSLENAFAFALNOMPALTYVANDGWEDINNTIAENALRAVSGRKN
FLFRGSDHGBRGALYSLICTKLNDVDPSYLRLHYLVGIADMPVARNVSELPRRIA
LPAP"
gene
52673..52915
/gene="Ecs1105"
52673..52915
CDS
/gene="Ecs1105"
/translation="probable terminase small subunit, similar to
C-terminal parts of terminase small subunits e.g.
[Bacteriophage N15] g1j2507082|sp|P31061|NOHA_ECOLI
percent identity 46 in 75 aa, GYG start, probably disrupted
by IS insertion"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative terminase small subunit"
/protein_id="BAB34528.1"
/db_xref="GI:13360565"
/translation="MSKGLIDTGFICFALSKLMAALSTLSDIPLSMQRFPLTPR
HLHLKTLIAKGANOCARAGDKLPDLDEYIRATTE"
gene
52887..54815
/gene="Ecs1106"
52887..54815
CDS
/gene="Ecs1106"
/translation="terminase large subunit, similar to terminase large
subunits e.g. [Bacteriophage 21]
g1j285157|sp|P36693|ITERL_BPP21 percent identity 91 in 637
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative terminase large subunit"
/protein_id="BAB34529.1"
/db_xref="GI:13360566"
/translation="MNISEQQLNNMNSAVTALPLRALPVTPEVMNQVYLPKES
SYGGEKMTLPFOIAIMNSMGNDIRYVNLKISARVGTAKMLGGGYFIEHKRSNL
LFOPTDAAEDFMKSHVATIRVPCDKLDSPLMGRKRONITLTKRSSGCVGMCIG
GAAKNYREKSVDYCYDELSPFVDYKESPPVYKESPPVYKESPPVYKESPPVYK
GTCOIEKANSAPMRYVPCPHGEQVYLPKESPPVYKESPPVYKESPPVYKESPPV
HGVYHQSSELDNSGNGWICENTGMMTRGLTFEAGDEKIDPPRSMHWTASPTP
TWQIVIVDMLDALDPLNGLKFTVNTLGEWEBAVEKGLDQVLMKDVARTAAVPAV
VYVLTAGIDSRNREMYVMAGAEAFIVDKIIIMGRGDEEELRLVDAINKKRV
HADGTEMTISKVCMDTGIDGEIYORSKHGVRLPVKAGSYKGVKVTPTKTRQ
RGVYICGDTAKELILAKRKADPPADEATYAIRPDDPEIPIQSDEAOQLVAEEL
VEKNEKMRLLMKNKRRNALDCLVAYALNAVSVORWOLDLAVLAKSSEETTRP
TKLELAALSGVNGYSR"
gene
55002..56594
/gene="Ecs1107"
55002..56594
CDS
/gene="Ecs1107"
/translation="portal protein, similar to portal proteins e.g. GP4
[  
[Bacteriophage P21] g1j549295|sp|P36272|VG04_BPP21 percent  
identity 98 in 530 aa"  
/codon_start=1  
/transl_table=1  
/evidence=not_experimental  
/product="putative portal protein"  
/protein_id="BAB34530.1"  
/db_xref="GI:13360567"  
/translation="MKRPVPLDVNGVPLRESLSYNGGAGGFGGOMAEMLPPAOSADA  
ALPALRLGNARADDVLVRNNGCIANAVANALHKKDHLVGHMFLISYRPNRMKGMRETAK  
SFVDEVEANSEYAEAGSGEIDVEKRTFTFIRGCVYAHFNEITVQVWDTETTO  
LFRYRKAVSPKRVDTPEHGMGNRLRAGVVDVDRGRAVAHICEDDPKSGSRMER  
IPRELPTGRPALMHIPEPEDEGOTRGANQFYSVMERLMDLSQATQLOSAIVAWA  
ATIESDIDTEKAFEXIAGAPQGDNDPLINLIEKFSYSDNNYNTLGGVKIPHLFPGD  
DLKIOTADSDNGFSALQOALLRYIAAGLVGSYBOLSDYSDKYSASASANSWRY  
FMGRKFTIARLALQMSVYMLEALIRGIIIRPPARPDFVQARSASAEWICGRMA  
IDGLAKEVESWRIEAGISTYKEPALMGEDYODIFRQVAVESARQKAGLSRPVIA  
QAYQOQIAESKRPSEETTPRET"  
gene  
56584..58089  
/gene="Ecs1108"  
56584..58089  
CDS  
/gene="Ecs1108"  
/translation="head-tail preconnector protein, similar to head-tail  
preconnector proteins e.g. GP5 [Bacteriophage P21]  
g1j549296|sp|P36273|VG05_BPP21 percent identity 97 in 501  
aa, GYG start"  
/codon_start=1  
/transl_table=1  
/evidence=not_experimental  
/product="putative head-tail preconnector protein"  
/protein_id="BAB34531.1"  
/db_xref="GI:13360568"  
/translation="MRMLSHIIIAAFNEPLLEPAVAFECALGRMGAAISLSPQ  
QOYOLDAPGLAETDEYVAGGKRPARYRVYVNGAVLPLTGTIVYHRIKGMSPFGMTG  
YDGIYACLOAMADSOVGRVGLDIDSPGCAAGFDCADMTYRLROKPPWALCNDPA  
CSAMULASACSRILVYOTSRITSGICWMSVSTAGHIAQAGVDITLITSGAHVNDG  
QFEALPAEVRONMOQRIIDARMEKREKVAKMTGTSVDVAVGTETEAEGVGGIDAGA  
DELVNASDAISVMAITALNSVNGVTMPQLTATEAANOQWMLGTCQKGRBOLA  
TMLAGOOQSVSEARAILAAAPQPVASPTSEADRIACCEANGREOLATLAMP  
MTVEKARPIILASPOADGPELRODINLIDPAKGEAQAQOLACPGMTVSARAVLA  
AGSKAEVVSASTVTLAFRIANISPAVGVGVPQTSADGADVKMLAMM"  
gene  
58126..58473  
/gene="Ecs1109"  
58126..58473  
CDS  
/gene="Ecs1109"  
/translation="head decoration protein, similar to head decoration  
proteins e.g. Gshp [Bacteriophage P21]  
g1j549437|sp|P36275|VSH_PBP21 percent identity 95 in 115  
aa"  
/codon_start=1  
/transl_table=1  
/evidence=not_experimental  
/product="putative head decoration protein"  
/protein_id="BAB34532.1"  
/db_xref="GI:13360569"  
/translation="MYTNITTEORAEVRIIFAGNDPAHTATGSSISSATPALTLPLMD  
EASGLVWDOCKAGSAIGIIVLPLEGTETVLTYKSGTFATEAIRMPSEVDEHKAN  
AFASALSHALP"  
gene  
58531..58797  
/gene="Ecs1110"  
58531..58797  
CDS  
/gene="Ecs1110"  
/translation="possible major head protein, similar to N-terminal  
parts of major head proteins e.g. GP7 [Bacteriophage P21]  
g1j547612|sp|P36270|HEAD_BPP21 percent identity 95 in 88  
aa"  
/codon_start=1  
/transl_table=1  
/evidence=not_experimental  
/product="putative major head protein"  
/protein_id="BAB34533.1"  
/db_xref="GI:13360570"  
/translation="WGLFTTRQLGYTEQKVFRAFLFLEFRRTVNFHEEVMADKI
```


gene
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase OrfA protein of insertion sequence IS629"
/protein_id="BAB34512.1"
/db_xref="GI:13360549"
/translation="MKNRFSEVROAIRWLVESODEYDSOMALICSIAPIGCTUTETLRVAVRQHERHEDTGGDGLTSARORLKELERENRELRNSNDILRQASVPAKAEERDLRMKK"
45341..45994
/gene="ECS1090"
45341..45994
/gene="ECS1090"
/note="probable transposase, similar to transposases e.g. [Escherichia coli plasmid p0157 insertion sequence IS629] g117443862|pir1||T00240 percent identity 98 in 220 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transposase OrfB protein of insertion sequence IS629"
/protein_id="BAB34513.1"
/db_xref="GI:13360550"
/translation="MMPLADLKREQYGWPVCSELHAPSTYYHOOGQRHHDRKSARSAAHDMDLKEIORYVDENHQYGVRRKVRQLREGIRVARCTVARLMAVMGLACVLGRKMARRRSGCIHSDSGSOYSLAYTERLKEAGLLASTGSTGSYDNMAESIINGLYKRAEVIHRSMKNRAVEELATLTJWDVWNRRRLGRGHTPAPEAKRAYVASIONDLLAA"
complement(46199..46513)
/gene="ECS1091"
complement(46199..46513)
/gene="ECS1091"
/note="probable transcriptional regulator, similar to Perc (Bfpw) [Escherichia coli] g111172431|sp|P43475|PERC_ECOLI percent identity 47 in 87 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator"
/protein_id="BAB34514.1"
/db_xref="GI:13360551"
/translation="MLDHVAECLEKGLYRPAAREMAKVNVQUSDQRRVAAQKRACELRKARRPVSPVNLTEIKQAVNRHLSELGMEFEERRVFRRYKOTGEONTSGNARSKRC"
complement(46979..47446)
/gene="ECS1093"
complement(46979..47446)
/gene="ECS1093"
/note="endopeptidase (host cell lysis), similar to endopeptidases e.g. [Bacteriophage VT2-Sai] g1158616591dbj|BA084330.1| percent identity 88 in 154 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative endopeptidase"
/protein_id="BAB34515.1"
/db_xref="GI:13360552"
/translation="MNRVLCYIIYLIVACGSLSIGLHHYRDHAITYEQDKKASELELANITDMQIRQRDVNALDAKISRELADARAENETLRADVAGRKRLRNATCGSGVHEATQTSVDNATGRLADTAERYDFILIRELIAMOKOLEGCAOEYIRTQOIP"
complement(47041..47226)
/gene="ECS1092"
complement(47041..47226)
/gene="ECS1092"
/note="lipoprotein R21 protein precursor, similar to R21 precursors e.g. [Bacteriophage 933W] g1145854519bj|AAD25453.1|Af125570.48 percent identity 98 in 61 aa. [Bacteriophage lambda] g11540738|pir1|JUN0750 percent identity 70 in 61 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative lipoprotein R21 protein precursor"

```
/db_xref="GI:13360537"
/translation="MTKMHDIHCOTRIPNSINHDTKYKICIFTPKNIKNNAITYILD
GNSANNYSIDLPYDALPNPVLVTLGYESNNLSIHRRAVDYDPDGENAIVDSKP
AMVFTGGSGSFPRELLTQIMPVYSTIAPNSSRGICGHSGAIFVDCIKNKCNFN
YYISAPSLMONERIITIKIDDISSEKTHSICILNGLNSLDVSASLTPKATKESV
LRNITTEKYSNFSIVQPELNNHQTFSALMNSIIHFSI"
gene
36960..37121
/gene="ECS1078"
36960..37121
/gene="ECS1078"
/note="unknown"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34501.1"
/db_xref="GI:13360538"
/translation="MSLIKNLVFSYPIYLAHYDLRIYVSKSTLLFYLSIHLLI
IVYIACIGN"
complement(37284..37421)
/gene="ECS1079"
complement(37284..37421)
/gene="ECS1079"
/note="unknown"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34502.1"
/db_xref="GI:13360539"
/translation="MLTGAFLYLPVMPHEADSLKHQOLFYLTSHYLANMKSVRKTDV
P"
gene
37466..37678
/gene="ECS1080"
37466..37678
/gene="ECS1080"
/note="probable prophage maintenance protein, similar to
Hok/Gef family e.g. Mokw [Bacteriophage 933W]
gi14584543|gb|AAD25481.1|AF125520_76 percent identity 92
in 65 aa"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="prophage maintenance protein"
/protein_id="BAB34503.1"
/db_xref="GI:13360540"
/translation="MNKSRSLASYPKGKEQAKQKQKAMLIALIVICLTVIYALVT
RKDCVAVRTGTEVAVFAYPEPE"
gene
37846..38124
/gene="ECS1081"
37846..38124
/gene="ECS1081"
/note="unknown, similar to QD1 [Bacteriophage N15]
gi12564084|gb|AAB81659.1| percent identity 31 in 64 aa"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34504.1"
/db_xref="GI:13360541"
/translation="MAHDTKLYNSDDSAVFASSRRCRCPHFASKSDMYOHPCTEQCAEW
LIQCYRRRGCEKVALSLDYRHWITISVLPYSENRPRPSRTFGQRIM"
gene
38198..39175
/gene="ECS1082"
38198..39175
/gene="ECS1082"
/note="unknown"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34505.1"
/db_xref="GI:13360542"

/translation="MSAFNGRIIVEPEPKSRMLPSGVVPAVHOPLAEDKSLPEFS
DERVIRAGGAGALSDMILRHVKSQOMPHGDYHSEYIYHRGTGAMVLCMHCDQLR
DOTSESLQLOONLAAMMIVIRBAMGJORELSLESWAYCNOVVALPEAVS
RRSLDLPEAKIRSVYRESDIIPGETATSLKORNTALPPTHQOONPROEKTVS
IAVDESPKSEFKRPRKRWNEKTRKWKTOPCACCKRPADPHHLIGHGGGGGK
SHDIETPLCRHNHMLHADPLAFEBKHSQVDLIFRLDHAFAVGIG"
gene
39188..39559
/gene="ECS1083"
39188..39559
/gene="ECS1083"
/note="probable crossover junction
endodeoxyribonuclease, similar to Gp67 [Bacteriophage HK97]
gi16901639|gb|AF31142.1| percent identity 60 in 113
aa, also similar to crossover junction
endodeoxyribonuclease Rus [Escherichia coli cryptic
prophage DUP12] gi12"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="putative crossover junction
endodeoxyribonuclease"
/protein_id="BAB34506.1"
/db_xref="GI:13360543"
/translation="WRIEFVLPPPTVNTYMRRGSTYFVSKAGERYRAVALIVRQC
RLKLSISGRLAIKIIEPPDKRRDLNIIKAPLDALTHAEYLIIDDEQFDEINIVRG
FVPGGRLGVKIVEIRGNDA"
gene
39549..39920
/gene="ECS1084"
39549..39920
/gene="ECS1084"
/note="probable antitermination protein, similar to
bacteriophage antitermination proteins e.g. Ybcq
[Escherichia coli cryptic prophage DUP12]
gi14584546|gb|AAD25444.1|AF125520_39 percent identity 77
in 124 aa"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="putative antitermination protein"
/protein_id="BAB34507.1"
/db_xref="GI:13360544"
/translation="WARDIOWLVLRGMAANNHEDVTPSIAGFKLIPTKYSRI
KCSDDAMITIGCMARLNKNQYLDLDYDYGMTFMALARKHRCSDGLGRLYP
AEGITIEGMALNVLDDMDR"
gene
40072..40890
/gene="ECS1085"
40072..40890
/gene="ECS1085"
/note="unknown, similar to Ypbd [Bacillus subtilis]
gi11730886|sp|P50730|YPBD_BACSU percent identity 30 in 126
aa"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="hypothetical membrane protein"
/protein_id="BAB34508.1"
/db_xref="GI:13360545"
/translation="WMIVLVSLSTLSMHKVVASLSLTVSVYLAVIDIMSVLEFY
ATIVFIIILKFNWKRYNAMAKSIVEGVLSAIALFHLMPGHNVVINSVVGQSI
PYTFVFNFDKALVPELLVCTSSLFKKKSEVSLMKKGASLSVLLILFLAVEGGI
KPEIHPEMLPEFIIANLFEVSLAESLSLRGIOSRLSEVSPVALVAALVAFGFY
YSGGLIVLFAFTLSGVVYGLSMMSGRIMVATLIFRGLNLCHLLEFTYPLKHN"
gene
41177..41416
/gene="ECS1086"
41177..41416
/gene="ECS1086"
/note="unknown, similar to hypothetical protein
[Bacteriophage p27] gi18346569|emb|CAB93762.1| percent
identity 97 in 49 aa"
/codon_start=1
/translation_table=11
/evidence-not_experimental
/product="hypothetical protein"
```


| | | | |
|------|---|------|--|
| gene | /product="hypothetical protein" /protein_id="BAB34481.1" /db_xref="GI:13360518" /translation="MNTFALVTLVYLVSGESELVTLGYLGSMECAAAAAEQRKIPGN CYVDKTTTHNNNEIPAGL" complement(28192..28380) /gene="ECS1059" complement(28192..28380) /note="ECS1059" /note="probable cell division inhibition, similar to <i>dicB</i> [<i>Escherichia coli</i>] gi 2507009 sp P09557 DIB3_ECOLI percent identity 54 in 62 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative division inhibition protein" /protein_id="BAB34482.1" /db_xref="GI:13360519" /translation="MEKLSNASTSELRFEIGVTGDKTFIEDAIKORKLEQDLNEY CIPSLARIKDLKGRKQ" complement(28566..28618) /gene="dicf1" /note="inhibitor of <i>FtsZ</i> " /product="dicf antisense RNA" complement(28566..28618) /gene="dicf1" 28647..28952 /gene="ECS1060" 28647..28952 /gene="ECS1060" /note="unknown, TTG start" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34483.1" /db_xref="GI:13360520" /translation="MSLKRNVNLTTRCPQGRKNPAILTAAPFGLDPEAHGRIKEN DSKSGELRRALSCYALTFQGYILSVCCRIICVARTTSSVLVINSALNG" 28954..29139 /gene="ECS1061" 28954..29139 /gene="ECS1061" /note="ECS1061" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34484.1" /db_xref="GI:13360521" /translation="MPNKPCCPACNALSGLEKGGYIFNCPHEVHFISKLDSITRP NOYQSDLNKKELNAARD" complement(29177..29329) /gene="ECS1062" complement(29177..29329) /gene="ECS1062" /note="ECS1062" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34485.1" /db_xref="GI:13360522" /translation="MKGTASADFTYSMMADGEPPELKITPGNHMPVITRAENW FNPVQO" complement(29326..29715) /gene="ECS1063" complement(29326..29715) /gene="ECS1063" /note="unknown" /codon_start=1 /transl_table=11 /evidence=not_experimental | gene | /product="hypothetical protein" /protein_id="BAB34486.1" /db_xref="GI:13360523" /translation="MEFKDLKEIOEIAHAHLQRUNEVELESATKRIIDNMRNVDA FTGLYSVSTNNQNTETAKRIASVWGPFVEEKYSKKEFWKTTKLQSNCHLRQS LLSMRKVIOMODYRNCSTKNSLS" complement(29727..29855) /gene="ECS1064" complement(29727..29855) /gene="ECS1064" complement(29727..29855) /note="ECS1064" /note="unknown, similar to <i>YdfB</i> [<i>Escherichia coli</i>] gi 140585 sp P29009 YDFB_ECOLI percent identity 100% in 41 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34487.1" /db_xref="GI:13360524" /translation="MDFTIMEKAVEEYFEGLAEGEALSFSEKQALSSAKSND" complement(29857..30012) /gene="ECS1065" complement(29857..30012) /gene="ECS1065" complement(29857..30012) /note="ECS1065" /note="unknown, identical to <i>YdfA</i> [<i>Escherichia coli</i>] gi 140584 sp P29008 YDFA_ECOLI percent identity 100 in 51 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34488.1" /db_xref="GI:13360525" /translation="MDTIDLGNNEISLVGVFPNQDGTFTAMTYTKSKTFENGARRW LERNSGE" complement(30070..30288) /gene="ECS1066" complement(30070..30288) /gene="ECS1066" /note="ECS1066" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34489.1" /db_xref="GI:13360526" /translation="MNNFTPSFGFEGFCQKQDLKTVDTFYVCRLHTPTTHAPQNA GQYFELPSSGQVSSQPEAYEHDRYQL" complement(30290..30577) /gene="ECS1067" complement(30290..30577) /gene="ECS1067" /note="ECS1067" /note="unknown, similar to hypothetical proteins e.g. <i>yacB</i> [Plasmid ColiB-P9] gi 4512441 dbj BA075090.1 similar to 35 in 92 aa" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34490.1" /db_xref="GI:13360527" /translation="MLPILMLPSARDLRLQITTYIAKENPPARRLRIRIETSVPLS EHPYLPPESESGAREIVTHPNYIILYVAASSIETVSTHSRQFPFSI" complement(30577..30768) /gene="ECS1068" complement(30577..30768) /gene="ECS1068" /note="ECS1068" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34491.1" |
|------|---|------|--|

/gene="Ecs1050"
/note="similar to YCCV_ECOLI g111787200 percent identity
100 in 122 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34473.1"
/db_xref="GI:13360510"
/translation="WMNFTLISKVKSREVTMIASKFGIGQOVRHSLGLGVVDID
PVSLSSEPSDELAAYNDELRAAPMYHVMEDDNGLPVHTYLAEQLSSELODEHEQEP
SMDELAQTIKROLOAPRLRN"
complement(21085..22188)
/gene="Ecs1051"
complement(21085..22188)
/gene="Ecs1051"
/note="similar to B0967_ECOLI g111787201 percent identity
99 in 367 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative oxidoreductase"
/protein_id="BAB34474.1"
/db_xref="GI:13360511"
/translation="MEGKASIGETIDIVDHQKMLARGAVSPAQIRARVWTFPSES
IDIAFESRLOAOAKMDMLAKDKGDSYRIAGSDELPGITIDRGNFLVLDLSA
GAEYORALISALOTLYPECAITRSDVAARRKESMETDGLVTGELPALPLEHG
MKLLVDIOHGKRTGYLDQROSLRATRRYENKRLNCFSTYGTGFAVSAALMGCSQV
SVDTSEALDIARONVELNKLIDLKSAREVDVFKLLRTYRDGEEKFVIMDPKPY
ENKSQLMGACRGYDINMLAIQILNEGILITFSCSGIMTSDLPQKITADAAIDAGR
VOFIEFROADHDVVIATYPEGYLKGFACRVM"
22370..22648
/gene="Ecs1052"
22370..22648
/gene="Ecs1052"
/note="similar to B0968_ECOLI g111787203 percent identity
100 in 92 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34475.1"
/db_xref="GI:13360512"
/translation="MSKVCITIAWYGRVQGVGFFRTTYEAKKRLGLTYKANLDDGSV
EVVACGEGGYEKLQMLKSGSPRSARVERVLEPHPSGELTDFRLR"
complement(22645..23031)
/gene="Ecs1053"
complement(22645..23031)
/gene="Ecs1053"
/note="similar to YCCV_ECOLI g111787204 percent identity
99 in 128 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative sulfite reductase"
/protein_id="BAB34476.1"
/db_xref="GI:13360513"
/translation="MRGFVFCYTAGSLIKVITIMLIFEKGEIETDEGYLKSSOMSE
PLAVVIAENGISLSPEHMEVRFVLEFNTSPAIRMVYKAMANKFGEEKNSRY
LYRLFPGPAQAKIKIAGLPKPKCI"
complement(23065..23724)
/gene="Ecs1054"
complement(23065..23724)
/gene="Ecs1054"
/note="similar to YCCA_ECOLI g111787205 percent identity
99 in 219 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative carrier/transport protein"
/protein_id="BAB34477.1"
/db_xref="GI:13360514"
/translation="MDRISSSHDRSTSLSTHKVLRNTFYLLSLTALFSAITASTV

gene
CDS
complement(24132..25151)
/gene="Ecs1055"
complement(24132..25151)
/gene="Ecs1055"
/note="probable integrases similar to Integrases e.g.
[Bacteriophage HK022] g11138560|sp|p16407|vIntf_BPH0
percent identity 27 in 321 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative integrase"
/protein_id="BAB34478.1"
/db_xref="GI:13360515"
/translation="MGRPRKKKKQNVLPVRVRSNGSYVYMKPEGSTRGICGRKRTS
VAKWQNTLEKAKLHNITVAKLHMHFMDSPAFTELPRYOKDROHOKALVWFGK
VLADNVKTEQVRIEMDKRGLESTQANHELASLSRYGWMGKRGYVKNPKGVKFS
LKARTVYITDQYAIYAIEAIPOLIAEISYLCAARGDVELKMODIMDKGIYIBQ
NKTGTQIKEMSPRLRTAIOLARVNSCTCYVINTTGGKGVIAKTTLNMMNOAKRA
EOKYGPFGCNFHDIKAKGISDYESSRDKOIFSGHKTENVGLIYDKRTKITPFLDL
LVSK"
25244..25396
/gene="Ecs1056"
25244..25396
/gene="Ecs1056"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34479.1"
/db_xref="GI:13360516"
/translation="MGSTLGPITNGMPCSLNMCCEPFYRSYFVLSISAGDRVSYMA
SRFTA"
complement(25439..27910)
/gene="Ecs1057"
complement(25439..27910)
/gene="Ecs1057"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34480.1"
/db_xref="GI:13360517"
/translation="MSKIFICAIIPDEQAIKEDSAVAVATAIEAGDERARRAKRFHMOF
LEOPPAODCAVKFTVCEDEKGIIPRALDSMDTBYMOENRDEDSASFVEPEPSDPM
NVNFDKLSPEYONAVLVKFDTCENTIVDMVISAQELQEDNATFGHIVELMKMPV
NAMYBELKLHIGWKKRCEPAGAKPELOEMRIWKRREGEREKGYTSVDLTTRA
RVNOONTENAEKIGAVTVAIRREYKOTPKTLDELACALMPGVVDGNTDGTTHRA
TNEYIDKDREDMKRTISAMKROPALQTDROTTEGLVREPIDHKDPAVLNKTISY
LTTKGVEHEETDOSTDALQPSAQAIPVATASDQKNEILVEAPSVEREQEPF
VFTDVGGEKYGKRAKFLGDLKALAGTEISKEEYFARKNGTYGLPQNVDTAEDSQ
SEPKVYTADEVKNTMQAANISOPADKILAAISREGEFEGISDPMDPKVMGIGIOSRQ
DONQPVKONPEAEKONSPTQONQPEQPEQPEPAVOOELKCYACCGTCGGCPCDQ
AYVKNATYLETFFEDENAEAKNPPEMEGTEHLAKENTSSDOYHASDNKTGETANL
IKVNGHEISSTSLRMLHMLTDLSTWKNPAPINTSIAGRFFPDATGEMPEPFSKTD
LETAGVIDRDTIKMWLKQSEAOSSILTDEITLDDALLQFREFIDNSGEFFVQWG
NGATFENVILRRSYEROGICPPMRYRTDRVRYTVAVLGLVMDFARTTIPFEGSRHNA
LHDARYOAKYVASIWMOKLPSQADF"
complement(28004..28195)
/gene="Ecs1058"
complement(28004..28195)
/gene="Ecs1058"
/note="unknown, similar to hypothetical proteins e.g. ydtd
[Escherichia coli] g11140587|sp|p29010|YDFD_ECOLI percent
identity 63 in 63 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental

```
CDS      complement(12716..13225)
/gene="Ecs1042"
/note="similar to SUDA_ECOLI g11787192 percent identity
100 in 169 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="suppressor of lon"
/protein_id="BAB34465.1"
/db_xref="GI:13360502"
/translacion="MTSGVHAHRSSSFSSASKIARVSTENTAGLISEVYREDPM
MTGOLLPLDQGLGQSRMGLMTLPQOKLSEWYAGSLPTKVMQISQSPCTFVS
MVRALRTGNSVIVIGMLADILTEEBHELVDANEGNMGFIMRVSASSSHATQISG
LKHISNDLYH"
13444..14073
/gene="Ecs1043"
13444..14073
/gene="Ecs1043"
/note="similar to B0959_ECOLI g11787193 percent identity
99 in 209 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34466.1"
/db_xref="GI:13360503"
/translacion="MKSLSYKRIYKSOEYATLGTIEYRSLSFGSYLTVDYPAVMS
DGELYLRACOSAGCYKVPYVLTFRKCGRSVTLYRYDESIMRQDLIVRSKXS
LDALKEKSTRNTRERKLDLPNMFHLEALIGEYGTIDVRLRLTGAKMKCLRQON
SLVTEKILFMELGALIGIHEALPVARROELAEWADSLTPKOEPAELE"
complement(14036..16198)
/gene="Ecs1044"
complement(14036..16198)
/gene="Ecs1044"
/note="similar to B0960_ECOLI g11787194 percent identity
99 in 720 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34467.1"
/db_xref="GI:13360504"
/translacion="MAFMLSPLKRYTMSAWLYARIFALCGTATPMMGLDVKLT
IPLTGNVAVALDLDLDLAGRLRNLIITLCEFIASASVELLPMPMLAIGLITST
SGFLLGIGORYATIAFGALLIATYMLGSLYEHWYQOPMYTLACAVWYNYLTIG
HLEPVRPLDNDLARCYEOLARYLELSKRMEDPDIEDESQAPLYDLALAGLIMATIN
OTKLSLTRLRGDRGCGRTRLHYVADIDHRASSSHIQYOTLEHRRHSVLER
FORLMSWQACQOOLSRICILROPYQDPHERAFTHIDALEMRDNGAPADILKTL
GFLNNLRADADLATTESQOALPHNNDENELADSPGLSDIMURLSRHFPESA
LFRAIVMSLVLCFGVAYIQTGMHGYWLLTSLFCQPNYATRHRLKRIITGLY
GIAIGIVLWFSVLEGOVLVITGVLPFAFRWQYAHATMTLTLLVLCFNLGEG
FEVALPREVIDTLIGCAITAMAASVYIMPDQPRNLRMLERATENCYDAITLBOYVO
GRONRLAVRIARDAHNRDAELASVSNMSEPRVTPQIRERARLLCLNHTTSYIS
ALGAHREQLNPELTLAFDDAVCTVDALHQPDEREYNVDEALSLKORQOLEPRAD
SKEPLVVOVGLLIALPEIGRLQRTQVPOETPVSA"
complement(16208..16654)
/gene="Ecs1045"
complement(16208..16654)
/gene="Ecs1045"
/note="similar to YCCF_ECOLI g11787195 percent identity
99 in 148 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34468.1"
/db_xref="GI:13360505"
/translacion="MRTVLNINLVLGCFATTLGMLATLVSIALLTFLPLFRSCWEI
TKLSLVYGENEALIHDELNPAGKNVLLNTGTVLNIPMLIFPGWMLGIMHAIAGIAC
ISIIIGIPVGTANRKAIALMLPVGRRVSVETQAAERANRRRFE"
16777..16831
/gene="Ecs1046"
16777..16831
CDS      complement(16777..16831)

gene      complement(18863..19330)
/gene="Ecs1046"
/note="similar to HEID_ECOLI g11787196 percent identity
99 in 684 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="DNA helicase IV"
/protein_id="BAB34469.1"
/db_xref="GI:13360506"
/translacion="MELKATTICKRLAONHPYDRAVILINAGIKVSGDRHEVILPENOLL
AIHKRCLVNGELFVLPDEKRYVRLHTEENGEDORFHHDAHRRRSGESSEIASV
LRQOLDLIARTIGENKMLITRQYSGVOQQLRQALSLPLPVNRLFEEDNCKEARQCO
AMLMDIESARQHNQAVTEAMLEYPDFRQVSSPLNPQARAIVNGESLSVLAA
GSGKTSYLVARAGVILARGASPEOIIILAFGRKAEMEERIERELHTEIDTARTER
ALALHIIOGSKKVPYISKLENDPAARHELFIEMRQCSSEKKAQAKGOMLITEEMQ
MSYPEGNFMDEKORLRLASRDWVSLMHHGGAQAEMLASAEFIRDLFSRKIKM
APLLKANKGALKANVDFSGLIHQAVILEKGFISPKKHILLVDEQDISPORAILL
AALKRQSSQTLLEAVGDQWQATYRESQAWSLTTFAEHENGEGERCUDITTYRNSRI
GEVANREIQONPGOLKRLNLTNGKRAVTLDESOLDLKLKSGYARPEERILTL
ARYHMRPASLEKATRWPKLQIDFMTIHASKQOADVILVIGLOESGDFPAAARS
IMEFALLPVEDFPDAERRLMYALTRARHRAWLFENKSPFVELKNLDVYVAR
Kp"
complement(19417..20079)
/gene="Ecs1047"
complement(19417..20079)
/gene="Ecs1047"
/note="similar to MESA_ECOLI g11787197 percent identity
100 in 155 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="methylglyoxal synthase"
/protein_id="BAB34470.1"
/db_xref="GI:13360507"
/translacion="MYTMEITTRLPRKRIHALYVNDHRCOMLSMYVERQPLEQNY
LYATGTGNLISRATGMNVNAMSGLPGQOYGALISEKIDVLTIFWQPLNAPHD
PVKALLRLATVNIPIVATVATADFIQSPHNDADVILIPYQRYLADRLK"
complement(20107..21078)
/gene="Ecs1048"
complement(20107..21078)
/gene="Ecs1048"
/note="similar to B0964_ECOLI g11787198 percent identity
100 in 220 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34471.1"
/db_xref="GI:13360508"
/translacion="MTGTIVTTLIALCLPYSVFATTLRLSTDVDLLVLDGKRVSSIL
RGADSTELDGPOLYERVEKTHLSSEERLISPLVVSFTOLINOVNPLRPLE
NEBEPANFPDAPELIDGATPIPVKIDILAITSAKTIDYEVYERVKSARRSL
PQRTAMADSTLSGSELDALIPQSOVLTTEORLKTWFKLADPQTRNTFLQNAEKOP
SS"
20252..20665
/gene="Ecs1049"
20252..20665
/gene="Ecs1049"
/note="similar to B0965_ECOLI g11787199 percent identity
99 in 141 aa, N terminus is different from B0965
(Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34472.1"
/db_xref="GI:13360509"
/translacion="MKETDIAGLITSTHTTALVGASDPRPSRYVMKYLLDGGYVVI
PSPKVAAGTKLIGQOGITLADYPERKVDVDRNSAAMGAVQAEALAIAGAKTLNMQL
GVINEQAAYIARPAGCLNVMDRCPALEIRLGLAK"
complement(20710..21078)
/gene="Ecs1050"
complement(20710..21078)
```

| | |
|--|---|
| gene | gene |
| CDS | CDS |
| 4307..5560 | complement(8258..8776) |
| /gene="Ecs1034" | /gene="Ecs1038" |
| /note="similar to PQA_ECOLI g11787183 percent identity 99 in 417 aa (Conserved in E.coli K-12)" | /note="similar to FABA_ECOLI g11787187 percent identity 100 in 172 aa (Conserved in E.coli K-12)" |
| /codon_start=1 | /codon_start=1 |
| /transl_table=11 | /transl_table=11 |
| /evidence=not experimental | /evidence=not experimental |
| /product="parquat-inducible protein A" | /product="beta-hydroxydecanoyl thioester dehydrase" |
| /protein_id="BAB34457.1" | /protein_id="BAB34461.1" |
| /db_xref="GI:13360494" | /db_xref="GI:13360498" |
| /translation="MCEHHNAKHILCSQCDMLVALPRLHEGKACPCRCGTTLVAM DAPQRPYALALFMLNLNLPFNMNVAGVTSFTLLEIPGLSESDYASLGRF FLFVQLVPAPCLITILLNRAELPYRKELARVLFOIKTMGAITFLAGVYSY KIMAYSIGSSSLPWCFLFCYDLRFPQCYDRMLMDIAPMELKQPLPGVTGIR OGRSCCCTAIDPADPCPCRCGKGYVRRNSLQMTLALVPSIMLYPANTLPII VTDLLGSKMPSLILAGVILMSEGSYPAAVIFLASIMVPLKMIATIMLCMDKKG KRSERMLIYEVFEGVMSMDIVFVIAVLSALVRMGMSIYPANGALMFALVIM TMSAMTFDPRLSMDRQSESEHES" | /translation="MVDKRESYTKEDILLASGRGELFGAKGPGQLPAPNMLMDRVKMT ETGNEPDKGYEALDNPDLMPFGCHPFGDPVAPPCGLGDAMQVLGFTLWGSGS KGRALGVEKVFQVLPYAKKVTYRIHFKTVNRILMGLADEVLVDRLLYTSAD LKVLFDQTSAP" |
| 5565..7205 | complement(8845..10605) |
| /gene="Ecs1035" | /gene="Ecs1039" |
| 5565..7205 | complement(8845..10605) |
| /gene="Ecs1035" | /gene="Ecs1039" |
| /note="similar to PQB_ECOLI g11787184 percent identity 99 in 546 aa (Conserved in E.coli K-12)" | /note="similar to B0955_ECOLI g11787188 percent identity 99 in 586 aa (Conserved in E.coli K-12)" |
| /codon_start=1 | /codon_start=1 |
| /transl_table=11 | /transl_table=11 |
| /evidence=not experimental | /evidence=not experimental |
| /product="parquat-inducible protein B" | /product="putative ATP-dependent protease" |
| /protein_id="BAB34458.1" | /protein_id="BAB34462.1" |
| /db_xref="GI:13360495" | /db_xref="GI:13360499" |
| /translation="MESNGEAKIOKYNKSPWPIPIVATLIGAWLTFHYHSGE VTLITANAEIGEGKTIKRSVDYGVESATLADDLITHEIKARLNSGHEKLIKOT VFNVVKQIRREGISGLTSLGVYIELPGAKSKMDKDLDSPLAPDPAKGIKIV VLBSKRAQVLSQDPDVLFRGYRVGSVETSTEDQKRNISYQLFNAPYDRLVSNVF WKDSGLAVDLTSGMRVEMGSLFTLLSGVSPDEGLDGPVAPKATAVLDVDDOKS IODSLYDHDILYLMFKDSYRGLQPCAPVEPRGIRGLTVSKVPEEPANMOTENDYR IPVLIRIEPRILKMLQGENADVVEHGLKRLGRISLKTGNLYTGALVYDDEPYPT PATIGTREFNGYQIIPYVSGGLAQIOQRLEALDKIKLELNPITEGATSTLESORT MKNLQITLDSMNKILASQSMQQLPTDMOSTLRELNRSQGFQPSAAVYNNKVAADMRL DQVRELQPYLTKLTNEKSNALVEFAKDKDPEPKRAQ" | /translation="MTITKLAMRLVPTDSYQEIFAQPHLIDENPLFSDTPRLOF ALBOLHTFRASSFMLAKAPESEYKLLINAAATLSDAGOLVGGHYEVSIRLR HAYSADNPFITQVNAADVVEAEOLFGCLROKAFDITDLPGLVHONGSILISLPT LIAQPLIMRLKINVNERFDWAFDSRPLPSPVSPMLKLVYLIGERESLADFE MERBELSQATYSEEDTLQIVDAESVQWCKRWTFETARHNLDPAGADAPVLIREA RYDEQETLPLSPQWILRQCKEVALSDGDTFSGEQNLMLQOREMREGLIARMODE LIOEQLILETEGRIQGINALSVLEFPHRATGEPSPRIQVYHIGDEPTDIERRAE LGNIIHAKGMIMQIAFMSELQLEQOIPFASLTFEESYSEVGDASMAELCALISA LADVPVOSIAITGSYQFRAQDPVGLNKGFFELICQORELTGQGVITFANVR HLSHLSVLVAEGRKTTIYAVDVPDALPULLNLVMDSGQITLMQTIQERIAQSQ OEGCHRPWPLRLNIMNFTPN" |
| 7217..7765 | 10791..11243 |
| /gene="Ecs1036" | /gene="Ecs1040" |
| 7217..7765 | 10791..11243 |
| /gene="Ecs1036" | /gene="Ecs1040" |
| /note="similar to YMB_A_ECOLI g11787185 percent identity 98 in 182 aa (Conserved in E.coli K-12)" | /note="similar to YCBG_ECOLI g11787190 percent identity 100 in 150 aa (Conserved in E.coli K-12)" |
| /codon_start=1 | /codon_start=1 |
| /transl_table=11 | /transl_table=11 |
| /evidence=not experimental | /evidence=not experimental |
| /product="hypothetical protein" | /product="putative dehydrogenase" |
| /protein_id="BAB34459.1" | /protein_id="BAB34463.1" |
| /db_xref="GI:13360496" | /db_xref="GI:13360500" |
| /translation="MTIALMLAGCSSGEINKNYQLPVQSGTOSTASQGNRLIWE QVAVPYLAGNGVYQTSQVYIANNNNLWASPLDRLMTLAVNSTOLRPGVAVASO PLGSADTLNVYTFENGRYDKVIYVSGEMLNHQGLIKRPRLESVQTIQDGTDEKV KYIAGVMSQFAAIAQETIKRLP" | /translation="MKYQLENLSESGMKKRYLVKKHREGELLTRYTEASAQEAVDVL LSLNEPVLVNGWIDKRMNPELVNRMKQITRARRKRHFNFNEHQTTRKSIDLEFIWQ RLAGLQORRGKTYSETIVOLIEDAENKEXYANKMSLSIKQDLALCKKE" |
| 8021..8188 | complement(11319..12359) |
| /gene="Ecs1037" | /gene="Ecs1041" |
| 8021..8188 | complement(11319..12359) |
| /gene="Ecs1037" | /gene="Ecs1041" |
| /note="identical to RMF_ECOLI g11787186 (Conserved in E.coli K-12)" | /note="similar to OMPA_ECOLI g11787191 percent identity 100 in 346 aa (Conserved in E.coli K-12)" |
| /codon_start=1 | /codon_start=1 |
| /transl_table=11 | /transl_table=11 |
| /evidence=not experimental | /evidence=not experimental |
| /product="outer membrane protein 3a" | /product="outer membrane protein 3a" |
| /protein_id="BAB34464.1" | /protein_id="BAB34464.1" |
| /db_xref="GI:13360501" | /db_xref="GI:13360501" |
| /translation="MKKTAIAIIVALAGFATVNOAPKQNTWTGAKLGSQYHDFG INNCGPHEHQACAGAPGQVNPYVGFPEBGYDMLRMKPGSVENGATKAGQVQLTA KICPTITDDIDITRIGGMYWRADTKSNVYGKHHGDSVPEFAGVEYATITPEIATRL ETQWTNNIGDAHTITGRPDNGMLSLGVSFRFGGEAPVAVAPAPAEVQTRKHFTLK SDVLFENKATLKEGEOAALDQLYSLNLDPRGDSVVVLYGYDRIGSDAYNOGLSER RAOVSVDYLSIKGIPADKISARGMGSENPVTGTCNVVQRAALLDCLAPDRRVEIEV KGIKIVVTOPOA" | /translation="MKKTAIAIIVALAGFATVNOAPKQNTWTGAKLGSQYHDFG INNCGPHEHQACAGAPGQVNPYVGFPEBGYDMLRMKPGSVENGATKAGQVQLTA KICPTITDDIDITRIGGMYWRADTKSNVYGKHHGDSVPEFAGVEYATITPEIATRL ETQWTNNIGDAHTITGRPDNGMLSLGVSFRFGGEAPVAVAPAPAEVQTRKHFTLK SDVLFENKATLKEGEOAALDQLYSLNLDPRGDSVVVLYGYDRIGSDAYNOGLSER RAOVSVDYLSIKGIPADKISARGMGSENPVTGTCNVVQRAALLDCLAPDRRVEIEV KGIKIVVTOPOA" |
| complement(8258..8776) | complement(12716..13225) |
| /gene="Ecs1038" | /gene="Ecs1042" |
| gene | gene |

186 GTTGTGCTGCTGTGTCGAGAGTGTGGCAATCAACGTTCTGCTGCTGTCGTGATTT
220 225
246 CGGTAACCGTGGTACCGAGACGATGCCGCCGGGCTGCAGTGGCCGTTCTCTCTGT
263 268
266 271
306 GCGCCCGCCGCGATCTGCGCGCGCGGTGACTGAGTACCTGCGCTGGACGACCGGACC
352 357
366 GCGACTTCATCGGATTCGCCGGGAAGTGTGATCTTGCCA
411 GGATGGCCGAGACGTCGCGTCATCTCCCGAGAGAGATCCGCTCAGCGGATATTA
461 466
471 C

3134 matches found in sequence:
ap002554 : TOIG of: ap002554 check: 6396 from: 1 to: 327773
(from "mycobacter-ge.seq")
TOIG of: ap002554 check: 6396 from: 1 to: 327773
LOCUS AP002554 327773 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION AP002554 BAO00007
VERSION AP002554.1 GI:13360491
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
AUTHORS 1 (sites)
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL
MEDLINE 20198780
REFERENCE
AUTHORS 2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL
MEDLINE 2057336
REFERENCE
AUTHORS 3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL
MEDLINE 20564182
REFERENCE
AUTHORS 4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,

Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
MEDLINE 21156231
REFERENCE
AUTHORS 5 (bases 1 to 327773)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: kenken@kenken.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/
Fax: 81-6-6879-2047)
COMMENT genome project
FEATURES
source
1. 327773
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
150..2258
/gene="ECS1032"
150..2258
/gene="ECS1032"
/note="similar to YCBY_ECOLI g111787181 percent identity
99 in 702 aa (conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="putative oxidoreductase"
/protein_id="BAB34455.1"
/db_xref="GI:13360492"
/translation="MNSLFASTARGLEBELKTELENLGAECVVOGVHFKGDRRLV
YSLWMSRLASRLMLPELGEKVSDDLVLGVQALNMTFNGAFVAFVFGSLNDTI
RNSOYGMKRVKDAIVDAEPFRKPLPRNVYRDADPIENVNMLHKEFASIALDSGLDI
LNGYRDRAIAPLKIETLAAIYVRSQWGPDLIDMCSSGTLIEAAMLATDRAGL
HGRMGFSGMAOHDEAIMEQVAKAETRRARKGLAEYSNFGSDSPARVIOARTAR
LAGIGLITFEYKDAQVQNLNPPPKPEYGVLSNPPGELDSERPAIIAHSLSLRMK
NOFGNNLSLSFNSPDLSCLOLRADKOKANKGDLVCQKNTHAESTVPSKPANA
EDYANLRNLRNKKFERKWARQEGIECYRLYDADLPENYNAVDRYADWVVOEYAPKTI
DAHKARQEDLIIAFTISVGIAPNKLVLKTEROKNQYKGLGKGFELVETENA
HLMVNLTVLDLIGLFDHRIARMLGMSKGFELFTYGSATYHAGLGARSTT
VMSRPTYLEMARNLRLNGLTGRARHLIOADCLAMREANBOFDLITFDPPNSKR
MEDAEDVQDRIALMKDLKRLRLRAGSTTFNSNKKGRFRDLGLAKLGKAEITOKT
LSQDFARNQRIHNCWLITAA"
2270..4177
/gene="ECS1033"
2270..4177
/gene="ECS1033"
/note="similar to UUP_ECOLI g11787182 percent identity 99
in 635 aa (conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="putative ATP-binding component of a transport
system"
/protein_id="BAB34456.1"
/db_xref="GI:13360493"
/translation="MSLISMHGAMLSFSDAPLLDNAELHIEDNERVCLVGRNAGKST
LKKILNRGSDGDRITVYEDLIVARLQDDPPRNVEGSYDVAEIEBOAYELKRYH
DISRLVMDPSKKNINELAKNOGOLDHNLKLEININVLQIGLDPVVALSISGCG
WLRKALGRALVSNPVLILDEPTNHLIDETITDMLDEGLIKTNGITIFISHRSFTIRN
MATRIYVDLRGLVTVPGNVDOYLLEKEBALVVEIQNAEFORKLAQOEVWIRQGIKA
RTRNMGSRVRLKAMRERGERREVYAKMOVEEASRSGLVFEHEDVCYVDFDQKOL
VDFSAOYVADGKIALICPNCGKTTLLKLMGLOQADSGSLVGVGKLEVAFTDQARA
ELDPKQTVMDNLDAEKGQEVWVGKPRHYGYODLFLPHKRAMPFVRLASGGERRLI
LARILPKSNLILIDEPNLDVETLELEELIDSYOGCVILVSHDRQFVMDTVTECW
ITREGGKIGRYVGGHDARQGOEYVALKQVAPKNEEPAKRAETVRRSSKUSYKL
QRELEQDLPDLLEAKLEALQTVADSAFSFQPHQDTOKVLADMAAAEQELEQAFER
WYELALKNGG"
4307..5560
gene

924 AGATGAGAAATACACAGACCGAGCATTCGTGGCCATCGCC

978 GTGGACGATACATCGATGATGAATCTTGAGGTCCGCGTCGTGACCCCAAGATTCA
1028 1033

1038 AGGCTTACCTGACGACGACATGACGGGATATACAAAGCCGAGCCCTGCGGCGATCA
1073 1078 1090 1095

1098 ACCAGCCGCGCCCTTGCGGTGACACCCGCTTGATATCTGCCGCGGTGAATTGGCCC

1158 CGCAGCCCGTAGTGTAGGACGCTATGATATGAAAGCCGACTGTTGAGTTGTGCGC
1175 1180 1212 1217

1218 GCGTCTTCGTGTGACGCGGTGCTGTACGGCGTGTGACCTGCATGTTCGCCACCGGT

1278 GGTGTGAGTGGGCTGCGACACACTGCGCTTACCGCGCGCATGGCGTGAATCGTC
1307 1312 1325 1330 1337

1338 GCCACCTCTTCCGGTTGTGGCCGCGGAT
1340

1 match found in sequence:
af009358 : TOIG of: af009358 check: 1421 from: 1 to: 204
(from "mycobacter-ge.seq")
TOIG of: af009358 check: 1421 from: 1 to: 204

LOCUS AF009358 204 bp DNA BCT 10-SEP-1997
DEFINITION Mycobacterium tuberculosis orfA2-898 gene, partial cds.
ACCESSION AF009358
VERSION AF009358.1 GI:2384664
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 204)
REFERENCE
AUTHORS Nano, F.E., Doran, J.L., Treit, J.D. and Moran, A.J.
TITLE Identification of Mycobacterium tuberculosis peptides that
stimulate immune human peripheral blood monocytes
JOURNAL Unpublished
2 (bases 1 to 204)
REFERENCE
AUTHORS Nano, F.E., Doran, J.L., Treit, J.D. and Moran, A.J.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1997) Biochemistry and Microbiology, University
of Victoria, Petch Building, Victoria, B.C. V8W, Canada
FEATURES
source
1..204
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
28..>204
/gene="orfA2-898"
28..>204
/gene="orfA2-898"
/codon_start=1
/transl_table=1
/protein_id="AAB69854.1"
/db_xref="GI:2384665"
/translation="MGVSPVRVRCVTGVALVALCMVLAGCTVSPPPAQSSDTPRS

BASE COUNT 29 a 65 c 75 g 35 t
ORIGIN
AF009358 Length: 204 February 21, 2002 16:58 Type: N Check: 1421 ..
Found using 'seq2-3' (pappu403.key)

131 CACCCAGAGCAGTGTATACGCCGCGACACACCAAGATCCAGCTCAGGCGATATTA
181 186

191 CTGCACCCGCGCGT

7 matches found in sequence:
af009359 : TOIG of: af009359 check: 8415 from: 1 to: 471
(from "mycobacter-ge.seq")
TOIG of: af009359 check: 8415 from: 1 to: 471

LOCUS AF009359 471 bp DNA BCT 10-SEP-1997
DEFINITION Mycobacterium tuberculosis orfH-200 gene, partial cds.
ACCESSION AF009359
VERSION AF009359.1 GI:2384666
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 471)
REFERENCE
AUTHORS Nano, F.E., Doran, J.L., Treit, J.D. and Moran, A.J.
TITLE Identification of Mycobacterium tuberculosis peptides that
stimulate immune human peripheral blood monocytes
JOURNAL Unpublished
2 (bases 1 to 471)
REFERENCE
AUTHORS Nano, F.E., Doran, J.L., Treit, J.D. and Moran, A.J.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1997) Biochemistry and Microbiology, University
of Victoria, Petch Building, Victoria, B.C. V8W, Canada
FEATURES
source
1..471
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
68..471
/gene="orfH-200"
68..>471
/gene="orfH-200"
/codon_start=1
/transl_table=1
/protein_id="AAB69855.1"
/db_xref="GI:2384667"
/translation="MAWRLVYFDAMHOLATIAIAIGVRLAPCGFRFRACWCM
WSKWOISFWLNRDVSIVGTDDAPGIRLAVFLCAAISAAYVTGYLRWTTDPDRF
NRDSREVYHATGMAETVVASPSPEGSNOGDIT"

BASE COUNT 65 a 132 c 167 g 107 t

AF009359 Length: 471 February 21, 2002 16:58 Type: N Check: 8415 ..
Found using 'seq2-3' (pappu403.key)

126 TGCCGCGATGCGGCTGTGCTGATGCTGCGACCAACTGCGCCACGATCGCGCC
137 142

> 0 <
01 10 Intelligenetics
> 0 <

Quest - Quick User-directed Expression Search Tool
Release 5.4

-- Outline of search "variant2-3" --

Selected search type is key against sequence data banks or files.

Selected scope is Sequence.

Selected sequence key from "pappu403.key":

```
1 seq2-3 (NA) ID seq2-3 NA preliminary pattern
2 either-or
3 followed by
4 a or g
4 a or g
4 c g
4 c or t
4 c or t
3 followed by
4 a or g
4 t c g
4 c or t
4 c or t
2 followed by
3 a or g
3 a or g
3 c g a
3 c or t
```

Selected files:

File : mycobacter_ge.seq

-- Output Parameters --

```
Format Options:
Nucleic acid code matching      Exact
Find non-matching hits only     No
Report key used                  Yes
Note position of hit             Yes
Display full annotations         Yes
Sequence context                 50
```

-- Run Parameters --

```
Run mode      Batch
Time to start comparison now
Notify at end of run No
```

16 matches found in sequence:

af009357 : TOIG of: af009357 check: 7028 from: 1 to: 1367
(from "mycobacter_ge.seq")

TOIG of: af009357 check: 7028 from: 1 to: 1367

```
LOCUS      AF009357      1367 bp      DNA      BCT      10-SEP-1997
DEFINITION Mycobacterium tuberculosis cytochrome c oxidase subunit II (coxII)
ACCESSION  AF009357
VERSION    AF009357.1 GI:2384662
KEYWORDS
SOURCE     Mycobacterium tuberculosis.
ORGANISM   Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE  1 (bases 1 to 1367)
            Nano,F.E., Doran,J.L., Treit,J.D. and Moran,A.J.
            Identification of Mycobacterium tuberculosis peptides that
```

```
JOURNAL      stimulate immune human peripheral blood monocytes
REFERENCE    2 (bases 1 to 1367)
AUTHORS      Nano,F.E., Doran,J.L., Treit,J.D. and Moran,A.J.
TITLE        Direct Submission
JOURNAL      Submitted (19-JUN-1997) Biochemistry and Microbiology, University
              of Victoria, Petch Building, Victoria, B.C. V8W, Canada
FEATURES
SOURCE
1. 1367
   /organism="Mycobacterium tuberculosis"
   /strain="H37Rv"
   /db_xref="taxon:1773"
   62. 1176
   /gene="coxII"
   62. 1174
   /gene="coxII"
   /note="corresponds to ORF11c of Sanger Sequencing Center
   cosmid MTCY190"
   /codon_start=1
   /transl_table=1
   /product="cytochrome c oxidase subunit II"
   /protein_id="AAB69853.1"
   /db_xref="GI:2384663"
   /translation="MIDGAPVTPRGRQLRLSQCRPQSGPARGRLALAML
   GALAVTVSGGSEGLGIMPEGITPEALHRELIGAVIASLAVGIWGLIFMSAV
   FHRKNTDTELPDQGYNMPLELVLPFLIVLFEYTVVOEKMLQKDEEVL
   DTSFQNMWKRQYRVNKGDTLYDGDAPRRKRAMSKPGKDKYGEELVPGVGLN
   TEDPTYLNEPKVEITGTETIPVILVBSGRIEFQMASADYIHAFWPEFLKRDVMP
   NPVANSYVNFQIEITRTGTAFVGHCBMCTTYSMADEYRVVTPNDKAYLQORD
   GNTVAELRALINOPPLAVTTHPEFDRRELAPQVPG"
```

```
BASE COUNT      259 a      396 c      425 g      287 t
ORIGIN
```

AF009357 Length: 1367 February 21, 2002 16:58 Type: N Check: 7028
Found using 'seq2-3' (pappu403.key)

```
...
23      TTCTGCCGCGACCGCGTCTATCCAGTCGCGCCGACGTTGTGATACAGAGCGCCACACA      |-----|
                                         73      78
83      GTGACACCTCGCGGCGGCGGCGTGTGCACAGCGTGTGCAGTGCGAGGCGCTCAGCGCGGC      |-----|
                                         102 107 111 116
143     TCCGGAGGCGCTGCCCGGTGCTTCGACAGCTGGCGGCTCGCAGCAAGTCTGGGGCGCATG      |-----|
                                         175 180
203     GCCGTCACCGTCAGTGGATGACAGCTGGT                                     |-----|
...
261     GCATTACCCCGGAGGACACACCTCAATGAGAACTGTGATCGGGCGGGTATCGGCCCC      |-----|
                                         311 316
321     TGGCGGTTGGGTAATCGTGTGGGGTCTCATCTTCTGTGGTCCGGGT
704     CGCGGGCTCAACACCGGAGACCGGACCTTAAGTGAATTTCGACAGAGGTGAGAGCTTGGGC      |-----|
                                         754 759
764     ACCAGCACGGAATTCGCGTGTGCTGTGCGCGTCCGGCAGACGTA
...
864     TCTTGTTCACGCTGACGTGATGCTTAACCCGGTGGCAACACACTGGGTCAACGCTTCC      |-----|
                                         914 919
```

```

XX 03-MAR-1995; 95US-0398633.
PR 07-JUN-1995; 95US-0487748.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Levinson DA;
XX
DR WPI; 2001-272703/28.
XX
XX New murine or human 200 genes and their corresponding polypeptides,
PT useful for treating or diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders, e.g. inflammatory diseases or allergies
PT
XX
XX Claim 1; Fig 17; 109pp; English.
XX
XX The present sequence is expressed at higher levels in T helper 1 (TH1)
CC cells than in T helper 2 (TH2) cells. The invention relates to an
CC isolated nucleic acid molecule, which comprises the full length murine
CC 200 gene or full length human 200 gene nucleotide sequence. The nucleic
CC acids are useful for treating or diagnosing immune disorders,
CC especially T helper lymphocyte-related disorders, e.g. inflammatory
CC diseases (e.g. Crohn's disease), multiple sclerosis, Grave's disease,
CC contact dermatitis, psoriasis, asthma and allergies, or certain viral
CC (e.g. HIV) or bacterial (e.g. tuberculosis) infections.
XX
XX Sequence 2710 BP; 789 A; 592 C; 644 G; 678 T; 7 other;
SQ
AAF82609 Length: 2710 March 5, 2002 14:18 Type: N Check: 1752
Found using 'seq2-3' (pappu403.key)
...
248 CTGATGAAGAATGTGACATATCAGAAATCCAGATACCAGTAAAGGCGATCTCA
|-----|
298 303
308 ACAAGGAGATGTCTCTGATCATAAAGAAATGTGACTCTGGATGA
...
553 AATGGAACAAAATTTCCACATGGCTGATGAATTAAGGACTCTGGAGAACGATCAGA
|-----|
603 608
613 ACTGCTATCCACATGGAGTGGAGTCTCTGCTGGGTGACCTGG
...
2201 GAGCTCTGCAGCAATCTATATAGTCAGCAAAATATCTTTGGAGGACAGTCGTCAACCA
|-----|
2251 2256
2261 AATTGATTTCCAAGCGGTGGACCTCAGTTTCATCTCTGGCTTACAGC
...
-----
4 matches found in sequence:
aaf82610 ; Murine TH1 specific 54 gene cDNA.
(from "mycobactering.seq")
TOIG of: aaf82610 check: 6440 from: 1 to: 1257
ID AAF82610 standard; cDNA; 1257 BP.
XX
AC AAF82610;
XX
XX 18-JUN-2001 (first entry)
XX
XX Murine TH1 specific 54 gene cDNA.
DE
XX Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;

```

```

KW anti-inflammatory; antiallergic; dermatological; antiviral;
KW antibacterial; T helper lymphocyte modulator; gene therapy;
KW TH specific gene; 200 gene; immune disorder; inflammation;
KW infection; 54 gene; cysteine protease; ss.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
PH CDS 22..1137
FT /*tag= a
FT /product= "54 gene product"
FT mat_peptide 397..1134
FT /*tag= b
XX
XX US6204371-B1.
XX
XX 20-MAR-2001.
PD
XX
XX 01-MAR-1996; 96US-0609583.
XX
XX 03-MAR-1995; 95US-0398633.
PR 07-JUN-1995; 95US-0487748.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX Levinson DA;
PI
XX
XX WPI; 2001-272703/28.
DR P-PSDB; AAB81512.
XX
XX New murine or human 200 genes and their corresponding polypeptides,
PT useful for treating or diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders, e.g. inflammatory diseases or allergies
PT
XX
XX Example; Fig 23; 109pp; English.
XX
XX The present sequence is expressed at higher levels in T helper 1 (TH1)
CC cells than in T helper 2 (TH2) cells. The invention relates to an
CC isolated nucleic acid molecule, which comprises the full length murine
CC 200 gene or full length human 200 gene nucleotide sequence. The nucleic
CC acids are useful for treating or diagnosing immune disorders,
CC especially T helper lymphocyte-related disorders, e.g. inflammatory
CC diseases (e.g. Crohn's disease), multiple sclerosis, Grave's disease,
CC contact dermatitis, psoriasis, asthma and allergies, or certain viral
CC (e.g. HIV) or bacterial (e.g. tuberculosis) infections. The present
CC sequence encodes a cysteine protease.
XX
XX Sequence 1257 BP; 323 A; 357 C; 323 G; 253 T; 1 other;
SQ
AAF82610 Length: 1257 March 5, 2002 14:18 Type: N Check: 6440
Found using 'seq2-3' (pappu403.key)
...
488 CTCGCGACAACATCCAGGCTCTGTGGCGCATCAACACCAGCAGTTTGTGGACGTCCTCTG
|-----|
538 543
548 TGCAGGAGCTGTGGACTGCGAACGCTGTGGAATGTTGCCAATGGTGGCTTCTGTGTGGG
|-----|
569 574
608 ACGCATATCTAACTGTC
...
912 CAAGGAGAAAGAGGCGCATGCAGACAGGACAGTCTTGTGCCATTCTCGAAACGTCGCCCA
|-----|
962 967
965 970

```

```
200 GTGGAATGGTATTACTCAGATACAAATGAAAGTATTCTCT
...
-----
8 matches found in sequence:
aaf82604 ; Murine TH inducible 10 gene cDNA.
(from "mycobacterng.seq")
TOIG of: aaf82604 check: 1121 from: 1 to: 2055

ID AAF82604 standard; cDNA; 2055 BP.
XX
AC AAF82604;
XX
DT 18-JUN-2001 (first entry)
XX
DE Murine TH inducible 10 gene cDNA.
XX
KW Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW anti-inflammatory; antiallergic; dermatological; antiviral;
KW antibacterial; T helper lymphocyte modulator; gene therapy;
KW TH inducible gene; 200 gene; immune disorder; inflammation;
KW infection; 10 gene; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT 496..1509
FT /*tag= a
FT /product= "10 gene product"
FT
XX
PN US6204371-B1.
XX
PD 20-MAR-2001.
XX
XX 01-MAR-1996; 96US-0609583.
XX
XX 03-MAR-1995; 95US-0398633.
XX 07-JUN-1995; 95US-0487748.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Levinson DA;
XX
XX WPI: 2001-272703/28.
XX P-PSDB; AAB81510.
XX
XX New murine or human 200 genes and their corresponding polypeptides,
XX useful for treating or diagnosing immune disorders, especially T helper
XX lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX Example: Fig 9; 109pp; English.
XX
XX The present sequence is a T helper (TH) cell inducible gene. The
XX invention relates to an isolated nucleic acid molecule, which comprises
XX the full length murine 200 gene or full length human 200 gene nucleotide
XX sequence. The nucleic acids are useful for treating or diagnosing immune
XX disorders, especially T helper lymphocyte-related disorders,
XX e.g. inflammatory diseases (e.g. Crohn's disease), multiple sclerosis,
XX Grave's disease, contact dermatitis, psoriasis, asthma and allergies, or
XX certain viral (e.g. HIV) or bacterial (e.g. tuberculosis) infections.
XX
XX Sequence 2055 BP; 623 A; 460 C; 415 G; 557 T; 0 other;
SQ

AAF82604 Length: 2055 March 5, 2002 14:18 Type: N Check: 1121 ..
Found using 'seq2-3' (pappu403.key)

...
448 TTCCCTGCTGACAGCCTAAGAAGTGATGTAACTGCCACTGTGAAGACCATGGCGATGAAC
|-----|
498 503
```

```
508 AGCATGTGCATTGAAGAGCAGCCGCACCTCGAACACTATTTGTTCCGGTGGTCTACATA
|-----|
527 532

568 ATTGTGTTTATAGTCAGCGTCCAGCAGCAACATCGGATCTTTATGCGTATCCTTTCTGCAA
|-----|
583 588

628 GCGAAGAAGGA
...

719 ACACTTGGAATAAAGACAACTGGACTTTCTCTCCACCTTGTCGCAAGAGAGGTTTCT
|-----|
769 774

779 TCACCTACATGAACTTTACAGCAGCAGCGGTTCCTCACTTGCAATTGCCCTGGACCGCT
|-----|
807 812

839 ATTTAGCAGTCGCTCTACCCCTCTGAAGTTTTTCCTTCTTAAGACGAGAAGATTGCGGTTTA
|-----|
847 852

899 TTAC
...

1220 CCCCCCTCCACGTGATGGTGCTCATCCGCTGCGTGTTCCTTCTTAAGACGAGAAGATTGCGGTTTA
|-----|
1270 1275

1280 ACAAGTCTGGATGGCAGACGTTTACGGTGTACAGAGTCACAGTAGCCCTGACAGAGTCTAA
|-----|
1296 1301

1340 ACTGTGTTGCCG
...

-----
3 matches found in sequence:
aaf82609 ; Murine TH1 specific 200 gene cDNA.
(from "mycobacterng.seq")
TOIG of: aaf82609 check: 1752 from: 1 to: 2710

ID AAF82609 standard; cDNA; 2710 BP.
XX
AC AAF82609;
XX
DT 18-JUN-2001 (first entry)
XX
DE Murine TH1 specific 200 gene cDNA.
XX
KW Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW anti-inflammatory; antiallergic; dermatological; antiviral;
KW antibacterial; T helper lymphocyte modulator; gene therapy;
KW TH specific gene; 200 gene; immune disorder; inflammation;
KW infection; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT 40..884
FT /*tag= a
FT /product= "200 gene product"
FT
XX
PN US6204371-B1.
XX
XX 20-MAR-2001.
XX
XX 01-MAR-1996; 96US-0609583.
XX
```


PT antigens -
XX Claim 1; Fig 1; 61pp; English.
XX The present sequence encodes the I01 antigen of Mycobacterium avium.
CC The specification describes a method for diagnosing and monitoring
CC paratuberculosis infection in a human or non-human animal. The method
CC comprises assessing the presence or concentration of a polypeptide
CC derived from paratuberculosis, a nucleic acid encoding the polypeptide,
CC antibodies directed to the polypeptide or one or more infection markers
CC which result from in vitro presentation of the polypeptide in a sample
CC from the human or non-human animal. The polypeptide is especially the
CC I01 or I02 antigen. I01 and I02 polypeptide, polynucleotides, and
CC antibodies are useful for assessing the presence or concentration of
CC paratuberculosis in a sample. They are useful for stimulating an immune
CC response against or for preventing infection by M. avium ssp
CC paratuberculosis, for diagnosing and monitoring M. avium ssp
CC paratuberculosis infection and for preventing paratuberculosis or
CC Crohn's disease in a human or non-human animal, especially a human
CC or a cow.
XX Sequence 671 BP; 125 A; 220 C; 202 G; 116 T; 8 other;
SQ
AAFP80135 Length: 671 March 5, 2002 14:18 Type: N Check: 698 ..
Found using 'seq2-3' (pappu403.key)
1 ATGCGCTCTGCTGACCATCGCGGATCAGTTCCCGCGCTACGAGCTGACCGGCTTGATCGCG
19 24
61 GCGGACCTGTCCAAGGTGCGAGCGCCAGCAGCGCGGTGACTACTTACCACCGTCAACCGC
61 66 79 84
121 GAGGACCAACGCGG
...
143 GCGTGGTGTCTTCTGCGCCCAAGGACTTCACCTTCGCTCTGCCCCACCGAGATCGCCACCT
193 198
203 TCGGCAAGCTCAACGACGAGTTCGAGGACCGCGACGCCAGGTGCTCGGCTCTCGATCG
214 219 235 240 250 255
263 ACACGAGTTCGTCCTCACTTCACTGGCGCGCCAGCAGGAGGA
...
326 CGATGCTCTCGGACATCAAGCGGCACTGAGCCTGGCCACCGGTGTCTTCAACGCCGACG
376 381
386 GCGTGGCCGACCGCGGCACCTTCATCGTCGACCCCGAAGAGATCCAGTTCGTCGCG
409 414
446 TCACCGGGGTTGCGTGGCGCGCACTGGCGCAAGGGTGACCGAGCTGCGGCTGCGGCTGC
469 474
506 AGTCCGACGAGCTGTGCGGCTGCACTGGCGCAAGGGTGACCGAGCTGAACGCCACCG
549 554 561
566 AACTGCTCAAGCGCTCTGCTTAATTCGGGCCCGCTGGGTGACAGGCTGATCAGGAGCGG
...
626 AGATGACGCTCANNNGGGGNTTTCAGTNAGATTCTTNNAGNACAA
631 636

3 matches found in sequence:
aaf82603; Murine TH2-specific gene 103 cDNA clone 103.1.
(from "mycobacterng.seq")
TOIG of: aaf82603 check: 2381 from: 1 to: 255
ID AAF82603 standard; cDNA; 255 BP.
XX
XX AAF82603;
AC
XX
XX 18-JUN-2001 (first entry)
XX
XX Murine TH2-specific gene 103 cDNA clone 103.1.
XX
XX Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW anti-inflammatory; antiallergic; dermatological; antiviral;
KW antibacterial; T helper lymphocyte modulator; gene therapy;
KW TH2-specific gene; 200 gene; immune disorder; inflammation;
KW infection; 103 gene; ss.
XX
XX Mus sp.
OS
XX
XX USG204371-B1.
XX
XX 20-MAR-2001.
XX
XX 01-MAR-1996; 96US-0609583.
XX
XX 03-MAR-1995; 95US-0398633.
PR 07-JUN-1995; 95US-0487748.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Levinson DA;
PI
XX
XX WPI; 2001-272703/28.
XX
XX New murine or human 200 genes and their corresponding polypeptides,
PT useful for treating or diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders, e.g. inflammatory diseases or allergies
PT
XX
XX Example; Fig 4A; 109pp; English.
XX
XX The present sequence is a T helper (TH)2-specific gene which was
CC obtained by subjecting RNA samples from TH1 and TH2 cells to
CC differential display techniques. The invention relates to an isolated
CC nucleic acid molecule, which comprises the full length murine 200 gene
CC or full length human 200 gene nucleotide sequence. The nucleic acids
CC are useful for treating or diagnosing immune disorders, especially T
CC helper lymphocyte-related disorders, e.g. inflammatory diseases (e.g.
CC Crohn's disease), multiple sclerosis, Grave's disease, contact
CC dermatitis, psoriasis, asthma and allergies, or certain viral (e.g. HIV)
CC or bacterial (e.g. tuberculosis) infections.
XX
XX Sequence 255 BP; 82 A; 53 C; 58 G; 62 T; 0 other;
SQ
AAFP82603 Length: 255 March 5, 2002 14:18 Type: N Check: 2381 ..
Found using 'seq2-3' (pappu403.key)
1 TTAGCGCCATTGCCATAGAGAGACCTCAGCCATCAATCACTAGCATGATGACAGA
3 8
...
80 TTGGCAATTCTGACACTTCCCATGTATTGACAGTTACGAGGCGAGTAATCGTCTGG
130 135
140 GGTCTGGAATGAGCGTTTAAATTGTGAGATGCCCCCAAGAGAGCGCTGACCTTATCCT
183 188

```
112 TTCCCCGACGCGGGGTACCGGGGTGGCCTGACGTACTCCGGCCCATCCGCTGGGACCGCC 162 167
172 TGC CGGTTCGGGAGATCAACGGCATGAAGACGAAGGCATGTTGGCCAAGCGTGC CGC 183 188
232 ATCGCGGAGCAGGTGCTCGGACCGGCTCTGCGGATCTCGCCGCCCGCACCGTTTCGGTC 220 225
292 GGCGAAGTCCGGCGCCCTCGCGGTCTTCTGGCGGGGATCTGATCCAGAACGGGCGGGTCTG 310 315
352 CGGGTTGAGGTCTCTCGGTGCCAGTGCCTGCGAGCGACGTCTGCGGCTGCTGGTATGCG 388 393
397 402
412 GCCCGGTAGGCTCTCTCGGTCCACACGTCACGACCGTGCCCGGGCGGATGATCTATC 421 426
437 442
472 AGTTCCGCCCTGCTCGGCAGCCCGCCGAGGGCAGCCAGTTCCGCTCCGGCGTCTGATCGGG 484 489
519 524
532 TTGGGTCCGTCCGGCCAGCACACGAGCATCCACCCGAGGTCGAGACAGGGTCCCGGACG 613 618
630 635
592 GTGCACATCTCCAGTCGATGAACGCGCGGAGCTCGGGGACGTCGCGGCGCAGCAGCACG 666 671
690 695
652 TTGTTTCAATGTCGCGTGCATGATCCCGGTTTCGGCTCGTCCGGCCTCGCGGAG 693 698
712 TCCAGCCAGTCCGGGAGCACATGACCGGGAACGACTCGGGCGCGGATCTGATCAG 745 750
772 CTCGGGAGCCGGTGCCCAAGCAGCCAGCTGGGAAGCACCGAGACCGCGCGGATGTG 794 799
832 CCGCGCAGCAGCGCCAGCGCTGTCACCCCGCGGACCGGGCCCGCGGACCGCGTCGGA 842 847
892 GTCGAC 1099 1104
1015 CGATGCGGNACACGACCAGCGAGCTGCAATCTGTCTCGGACCCCTGGGGTCAACG 1065 1070
1075 GCGGTGCTGCTCCGCAACCGCGCGGATGTCGCGCGCGCTGCGGCGGCTCTCCA 1099 1104
1135 TGCGCGTTCGTTCACTGCTGCTCGGGTGGTGTCTTCGGAACGGGCGCGCGCCCGT 1150 1155
1193
1195 CGTCCGTCCGATACGGGATCTATCAGCAGGTAGGCGCTCCAGCCGCTACTCTTCGCCCA 1198
1255 GAACAGCGGTGCGCGCGGAGACGAGCGGTCTCTGCGCCAGATACACCGAGCGGT
```

```
1315 GGCGGCAT 1268 1273
...
1354 GTGAGATCGGATCTATCGCGCGCTGTGCGGAAAGCAGGAGCGCTAGGCGCTTGAC 1404 1409
1414 TAGCTCCCGCGGTGGCGAAGACGAGACTACATCGACCGAGCTTGCACAACATCGGG 1417 1422
1474 CCGTATCTGCCAGCTGAGGTTCGCGCTCTCGTCCGATCTATCGCGCGCACCGGCCGCTG 1513 1518
1534 CCCGGCACCGGTGGATCTGTCAGTACCCGGAAGCTCTTGAGAGCTAAGGCCAATTGG 1549 1554
1594 GAAGATACTTGGACCTTCCCATCAATAGAGGAAAGCATCGCCCTAGGGATCCGTAGCG 1631 1636
1654 GGCCCGGTGTTTCGAGTGAACCTTGGGCGAGGCA
...
13 matches found in sequence:
aaf80135 ; Nucleotide sequence of the I01 antigen.
(from "mycobacterng.seq")
TOIG of: aaf80135 check: 698 from: 1 to: 671
ID AAF80135 standard; DNA; 671 BP.
XX
AC AAF80135;
XX
DT 11-JUN-2001 (first entry)
XX
DE Nucleotide sequence of the I01 antigen.
XX
KW I01 antigen; paratuberculosis; I02 antigen; Crohn's disease; vaccine; ss.
XX
OS Mycobacterium avium.
XX
FH Key Location/Qualifiers
FT CDS 1..588
FT /tag= a
FT /product= "I01 antigen"
FT /transl_except= "(pos: 574..576, aa: Cys)"
XX
PN WO200118212-A2.
XX
PD 15-MAR-2001.
XX
PF 06-SEP-2000; 2000WO-GH03430.
XX
PR 06-SEP-1999; 99NO-0004317.
XX
PA (VESO-) VESO AS.
PA (JONE/) JONES E J.
XX
PI Olsen I, Wiker H;
XX
DR WPI: 2001-244574/25.
DR P-PSDB; AAB67761.
XX
PT Diagnosing or monitoring infection by Mycobacterium avium ssp.
PT paratuberculosis, by assessing presence or concentration of the
PT bacterial derived antigens or antibodies cross-reacting with the
```

592 ACATGCCACCGAAGGACGATCGGCGCGGGATCTGATCGGCAGGCATCAGAACAGTA
607 612
652 AGCGGTGTTCC
...
694 CGATGCCGAACACCGACACCGACGAGCAGCTGCAATCTCTCGGAGCCCTGGCGTCACGC
744 749
754 GCGCTGCTGGCTCCGCAACCGCGCGGATGTCGCGCGCGCGCTGCGGCGGCTCTCCA
754 759
814 TGGCGGTTCGTTTCAGTCGCTCGTCGCGTGTCTTCTGCAAGCGCGCGCGCGCT
829 834
874 CGTCCGTCGATACCGGATCTGATCAGCTCGGGAGCGGTCGCCAGCAACGCGAGCGT
877 923 928
934 GGGAGCACCGACGCGCGGATGTCGCCCGCGCAGCGCGCGTGCACCCCGCG
971 976
994 GGACCGGCGCGCGGACCGCGTGGAGTCGAC
...
1080 CACCGAGTAGCGCGTCCAGCGGTACTCTTGGCCCCAGAACAGCGGTGCGCTCGCGCGC
1130 1135
1140 AGACCGCGTCTCGCGCGCAGATACACCCAGCGGTGGCGGCAT
...
1216 GTGAGATCGGATCTATCGCGCGCTGTGCGGGAAGCAGCGCGTAGCGGCTGTCAC
1266 1271
1276 TACGTCGCGCGGTGCGGAGAGCAGGACTACATCGACCGGCTTGCAGCATCGG
1279 1284
...
1382 AACCGCGCGTCTCGCGGTGAGTCTCTCGTCCCGAGTCCCGTGCAGCGGAGGTGTCG
1432 1437
1435 1440
1442 GCCTGATGCGCGCGCGGTAGCGGTCTCTCGTCCAGCAACGCTCAGACCGTGCCTCGG
1446 1465 1470 1481 1486
1502 CGGATGATCTATCGCGCGCACCGCGCGGTGCGCGCGCGGTGCGGTGATGCTGTCGTCAGT
1513 1518
1519 1554
1562 ACCGAAGCTCTTGAGAGCTAAGGCCAATGGAAGATATCTTGGACCTTCCCATCATAG
...
1622 AGGAAGCATCCCGCTAGGGATCCGTAGCGGGCGCGGTGTTTTCAGTGAACCTGGGCA
1631 1636
1682 GGCA

...
31 matches found in sequence:
aaf25122 ; Nucleotide sequence of an 8-mer multi-epitope designated ME/D.
(from "mycobacterng.seq")
TOIG of: aaf25122 check: 3543 from: 1 to: 1749
ID AAF25122 standard; DNA; 1749 BP.
XX AC
XX AAF25122;
XX DT 30-APR-2001 (first entry)
XX DE
XX Nucleotide sequence of an 8-mer multi-epitope designated ME/D.
XX KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX OS Synthetic.
XX OS Mycobacterium vaccae.
XX FH Key Location/Qualifiers
XX ET CDS 1..1749
XX FT /*tag= a
XX FT /product= "multi-epitope"
XX PN W0200104140-A1.
XX PD 18-JAN-2001.
XX PF 10-JUL-2000; 2000WO-N200121.
XX PR 12-JUL-1999; 99US-0351348.
XX PR 29-NOV-1999; 99US-0450072.
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX PI Delcayre A;
XX DR WPI: 2001-168411/17.
XX DR P-PSDB; AAB31660.
XX PT Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX PT vaccae, useful for treating mycobacterial infections, immune disorders
XX PT and cancers
XX PS Claim 7; Page 68; 80pp; English.
XX CC The specification describes an immunogenic epitope of a Mycobacterium
XX CC vaccae antigen. The epitope is a stimulator of cytokine production.
XX CC The epitopes are useful for the treatment of immune disorders,
XX CC infectious diseases, especially tuberculosis, and cancer. They are
XX CC also useful for treatment of other mycobacterial infections such as
XX CC those caused by Mycobacterium avium. The epitopes are especially useful
XX CC for inducing TH1 immune responses, and for producing vaccines. The
XX CC present sequence encodes an 8-mer multi-epitope designated ME/D.
XX SQ Sequence 1749 BP; 275 A; 612 C; 581 G; 281 T; 0 other;
AAF25122 Length: 1749 March 5, 2002 14:18 Type: N Check: 3543 ..
Found using 'seq2-3' (pappu403.key)

1 ATGGCTACAGGCTCCCGGAGCTCCCTGCTCTGCTGGCTTTTGGCGCTGCTCTGCTGCGCTGG
18 23

61 CTTCAAGAGGGCA

...
|-----|


```
OS Mycobacterium vaccae.
XX WO200104140-A1.
XX 18-JAN-2001.
XX 10-JUL-2000; 2000WO-NZ00121.
XX 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Delcayre A;
XX WPI; 2001-168411/17.
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX vaccae, useful for treating mycobacterial infections, immune disorders
XX and cancers -
XX Example 4; Page 64; 80pp; English.
XX The specification describes an immunogenic epitope of a Mycobacterium
XX vaccae antigen. The epitope is a stimulator of cytokine production.
XX The epitopes are useful for the treatment of immune disorders,
XX infectious diseases, especially tuberculosis, and cancer. They are
XX also useful for treatment of other mycobacterial infections such as
XX those caused by Mycobacterium avium. The epitopes are especially useful
XX for inducing Th1 immune responses, and for producing vaccines. PCR
XX primers AAF25106-19 were used to amplify DNA encoding M. vaccae epitopes
XX of the invention.
XX Sequence 37 BP; 8 A; 12 C; 12 G; 5 T; 0 other;
XX
XX AAF25107 Length: 37 March 5, 2002 14:18 Type: N Check: 93 ..
XX Found using 'seq2-3' (pappu403.key)
1 AAGGAGGAGGATCCGCGCGCGAGTCTCCGTC
|-----|
26 31
1 match found in sequence:
aaf25109 ; PCR primer for DNA encoding a M. vaccae immunogenic epitope.
(from "mycobacterng.seq")
TOIG of: aaf25109 check: 9067 from: 1 to: 37
ID AAF25109 standard; DNA; 37 BP.
XX
XX AAF25109;
XX
XX 30-APR-2001 (first entry)
XX
XX PCR primer for DNA encoding a M. vaccae immunogenic epitope.
XX
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX cancer; mycobacterial infection; Th1 immune response; vaccine;
XX PCR primer; ss.
XX
XX Mycobacterium vaccae.
XX
XX WO200104140-A1.
XX
XX 18-JAN-2001.
XX
XX 10-JUL-2000; 2000WO-NZ00121.
XX
XX 12-JUL-1999; 99US-0351348.
XX
XX 29-NOV-1999; 99US-0450072.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Delcayre A;
XX
XX WPI; 2001-168411/17.
XX
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX vaccae, useful for treating mycobacterial infections, immune disorders
XX and cancers -
XX Example 4; Page 64; 80pp; English.
XX The specification describes an immunogenic epitope of a Mycobacterium
XX vaccae antigen. The epitope is a stimulator of cytokine production.
XX The epitopes are useful for the treatment of immune disorders,
XX infectious diseases, especially tuberculosis, and cancer. They are
XX also useful for treatment of other mycobacterial infections such as
XX those caused by Mycobacterium avium. The epitopes are especially useful
XX for inducing Th1 immune responses, and for producing vaccines. PCR
XX primers AAF25106-19 were used to amplify DNA encoding M. vaccae epitopes
XX of the invention.
XX Sequence 37 BP; 8 A; 12 C; 12 G; 5 T; 0 other;
XX
```

```
PI Delcayre A;
XX
XX WPI; 2001-168411/17.
XX
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX vaccae, useful for treating mycobacterial infections, immune disorders
XX and cancers -
XX Example 4; Page 64; 80pp; English.
XX The specification describes an immunogenic epitope of a Mycobacterium
XX vaccae antigen. The epitope is a stimulator of cytokine production.
XX The epitopes are useful for the treatment of immune disorders,
XX infectious diseases, especially tuberculosis, and cancer. They are
XX also useful for treatment of other mycobacterial infections such as
XX those caused by Mycobacterium avium. The epitopes are especially useful
XX for inducing Th1 immune responses, and for producing vaccines. PCR
XX primers AAF25106-19 were used to amplify DNA encoding M. vaccae epitopes
XX of the invention.
XX Sequence 37 BP; 11 A; 7 C; 16 G; 3 T; 0 other;
XX
XX AAF25109 Length: 37 March 5, 2002 14:18 Type: N Check: 9067 ..
XX Found using 'seq2-3' (pappu403.key)
1 AAGGAGGAGGATCCGTCGCGAGCGCGGGG
|-----|
28 33
1 match found in sequence:
aaf25114 ; PCR primer for DNA encoding a M. vaccae immunogenic epitope.
(from "mycobacterng.seq")
TOIG of: aaf25114 check: 7762 from: 1 to: 22
ID AAF25114 standard; DNA; 22 BP.
XX
XX AAF25114;
XX
XX 30-APR-2001 (first entry)
XX
XX PCR primer for DNA encoding a M. vaccae immunogenic epitope.
XX
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX cancer; mycobacterial infection; Th1 immune response; vaccine;
XX PCR primer; ss.
XX
XX Mycobacterium vaccae.
XX
XX WO200104140-A1.
XX
XX 18-JAN-2001.
XX
XX 10-JUL-2000; 2000WO-NZ00121.
XX
XX 12-JUL-1999; 99US-0351348.
XX
XX 29-NOV-1999; 99US-0450072.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Delcayre A;
XX
XX WPI; 2001-168411/17.
XX
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX vaccae, useful for treating mycobacterial infections, immune disorders
XX and cancers -
XX Example 4; Page 65; 80pp; English.
XX The specification describes an immunogenic epitope of a Mycobacterium
XX vaccae antigen. The epitope is a stimulator of cytokine production.
XX The epitopes are useful for the treatment of immune disorders,
XX infectious diseases, especially tuberculosis, and cancer. They are
XX also useful for treatment of other mycobacterial infections such as
XX those caused by Mycobacterium avium. The epitopes are especially useful
XX for inducing Th1 immune responses, and for producing vaccines. PCR
XX primers AAF25106-19 were used to amplify DNA encoding M. vaccae epitopes
XX of the invention.
XX Sequence 37 BP; 11 A; 7 C; 16 G; 3 T; 0 other;
XX
```

```

1  AAGCTTATGGCTACAGGCTCCGCGACGTCCTGCTGCTCTGGGCTTTGGCCTGCTGCTGCTG
    24 29
61  CCCTGGCTTCAAGAGGGCA
...
-----
1 match found in sequence:
aaf25105 ; PCR primer for M. vaccae immunogenic epitope DNA9A.
(from "mycobacterieng.seq")
TOIG of: aaf25105 check: 7953 from: 1 to: 37

ID AAF25105 standard; DNA; 37 BP.
XX
AC AAF25105;
XX
DT 30-APR-2001 (first entry)
XX
DE PCR primer for M. vaccae immunogenic epitope DNA9A.
XX
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine;
KW PCR primer; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO200104140-A1.
XX
PD 18-JAN-2001.
XX
PF 10-JUL-2000; 2000WO-NZ00121.
XX
PR 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
PA Delcayre A;
XX
PI WPI; 2001-168411/17.
XX
DR Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -
XX
PS Example 2; Page 64; 80pp; English.
XX
CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders,
CC infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. PCR
CC primers AAF25104-05 were used to amplify DNA encoding a M. vaccae epitope
CC of the invention.
XX
SQ Sequence 37 BP; 12 A; 11 C; 13 G; 1 T; 0 other;

AAAF25105 Length: 37 March 5, 2002 14:18 Type: N Check: 7953
Found using 'seq2-3' (pappu403.key)

1  AAGGAGGAAGGATCCGCGCCAGAGCGCGAGGCC
    26 31
-----
1 match found in sequence:
aaf25106 ; PCR primer for DNA encoding a M. vaccae immunogenic epitope.
(from "mycobacterieng.seq")
TOIG of: aaf25106 check: 2746 from: 1 to: 34

```

```

ID AAF25106 standard; DNA; 34 BP.
XX
AC AAF25106;
XX
DT 30-APR-2001 (first entry)
XX
DE PCR primer for DNA encoding a M. vaccae immunogenic epitope.
XX
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine;
KW PCR primer; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO200104140-A1.
XX
PD 18-JAN-2001.
XX
PF 10-JUL-2000; 2000WO-NZ00121.
XX
PR 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
PA Delcayre A;
XX
PI WPI; 2001-168411/17.
XX
DR Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -
XX
PS Example 4; Page 64; 80pp; English.
XX
CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders,
CC infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. PCR
CC primers AAF25106-19 were used to amplify DNA encoding M. vaccae epitopes
CC of the invention.
XX
SQ Sequence 34 BP; 7 A; 9 C; 11 G; 7 T; 0 other;

AAAF25106 Length: 34 March 5, 2002 14:18 Type: N Check: 2746
Found using 'seq2-3' (pappu403.key)

1  GAGAGAGAGATCTATCAGTTCGGCCGCTGGCGCC
    29 34
    |-----|
-----
1 match found in sequence:
aaf25107 ; PCR primer for DNA encoding a M. vaccae immunogenic epitope.
(from "mycobacterieng.seq")
TOIG of: aaf25107 check: 93 from: 1 to: 37

ID AAF25107 standard; DNA; 37 BP.
XX
AC AAF25107;
XX
DT 30-APR-2001 (first entry)
XX
DE PCR primer for DNA encoding a M. vaccae immunogenic epitope.
XX
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine;
KW PCR primer; ss.
XX

```

CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. PCR
CC primers AAF25092-AAF25101 were used to amplify a construct comprising the
CC human growth hormone (hGH) signal peptide DNA fused to DNA encoding a
CC M. vaccae epitope of the invention. PCR primer AAF25092 is specific to DNA
CC encoding hGH, and PCR primers AAF25089-AAF25101 are specific to DNA
CC encoding M. vaccae epitopes.
XX

SQ Sequence 34 BP; 6 A; 11 C; 11 G; 6 T; 0 other;

AAAF25094 Length: 34 March 5, 2002 14:18 Type: N Check: 2709 ..
Found using 'seq2-3' (pappu403.key)

1 GAGAGAGAGATCTCGCGCCGAGTCGTCCCGTC
|-----|
23 28

1 match found in sequence:
aaf25100 ; PCR primer used to amplify hGH signal peptide-M. vaccae epitope DNA
(from "mycobacterng.seq")
TOIG of: aaf25100 check: 6452 from: 1 to: 36

ID AAF25100 standard; DNA: 36 BP.
XX
AC AAF25100;
XX
DT 30-APR-2001 (first entry)
XX
DE PCR primer used to amplify hGH signal peptide-M. vaccae epitope DNA.
XX
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine;
KW human growth hormone; hGH; PCR primer; ss.
XX

OS Mycobacterium vaccae.

XX WO200104140-A1.

XX 18-JAN-2001.

PF 10-JUL-2000; 2000WO-NZ00121.

XX 12-JUL-1999; 99US-0351348.

PR 29-NOV-1999; 99US-0450072.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA Delcayre A;

XX WPI; 2001-168411/17.

XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -

PS Example 2; Page 63; 80pp; English.

XX The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. PCR
CC primers AAF25092-AAF25101 were used to amplify a construct comprising the
CC human growth hormone (hGH) signal peptide DNA fused to DNA encoding a
CC M. vaccae epitope of the invention. PCR primer AAF25092 is specific to

CC DNA encoding hGH, and PCR primers AAF25089-AAF25101 are specific to DNA
CC encoding M. vaccae epitopes.

SQ Sequence 36 BP; 10 A; 7 C; 15 G; 4 T; 0 other;

AAAF25100 Length: 36 March 5, 2002 14:18 Type: N Check: 6452 ..
Found using 'seq2-3' (pappu403.key)

1 GAGAGAGAGATCTCGTATCGGACGCGGACG
|-----|
25 30

1 match found in sequence:
aaf25103 ; Nucleotide sequence of human growth hormone signal peptide.
(from "mycobacterng.seq")
TOIG of: aaf25103 check: 3442 from: 1 to: 126

ID AAF25103 standard; DNA: 126 BP.

XX AAF25103;

XX 30-APR-2001 (first entry)

XX Nucleotide sequence of human growth hormone signal peptide.

DE Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine;
KW human growth hormone; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 7..126

FT /*tag= a

XX WO200104140-A1.

XX 18-JAN-2001.

PF 10-JUL-2000; 2000WO-NZ00121.

XX 12-JUL-1999; 99US-0351348.

PR 29-NOV-1999; 99US-0450072.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA Delcayre A;

XX WPI; 2001-168411/17.

XX P-PSDB; AAB31657.

XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -

PS Example 2; Page 63; 80pp; English.

XX The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes the signal peptide of human growth hormone.
CC The signal peptide was used to express antigens of M. vaccae.

SQ Sequence 126 BP; 21 A; 48 C; 27 G; 30 T; 0 other;

AAAF25103 Length: 126 March 5, 2002 14:18 Type: N Check: 3442 ..
Found using 'seq2-3' (pappu403.key)


```
XX PF 10-JUL-2000; 2000WO-NZ00121.
XX PR 12-JUL-1999; 99US-0351348.
XX PR 29-NOV-1999; 99US-0450072.
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX PI Delcayre A;
XX XX WPI; 2001-168411/17.
XX DR Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX PT vaccae, useful for treating mycobacterial infections, immune disorders
XX PT and cancers -
XX PS Claim 4; Page 60; 80pp; English.
XX XX The specification describes an immunogenic epitope of a Mycobacterium
XX CC vaccae antigen. The epitope is a stimulator of cytokine production.
XX CC The epitopes are useful for the treatment of immune disorders,
XX CC infectious diseases, especially tuberculosis, and cancer. They are
XX CC also useful for treatment of other mycobacterial infections such as
XX CC those caused by Mycobacterium avium. The epitopes are especially useful
XX CC for inducing TH1 immune responses, and for producing vaccines. The
XX CC present sequence encodes a M. vaccae epitope of the invention.
XX SQ Sequence 177 BP; 27 A; 71 C; 66 G; 13 T; 0 other;
AAF25087 Length: 177 March 5, 2002 14:18 Type: N Check: 4115
Found using 'seq2-3' (pappu403.key)
1 GATCAGCTCGGGAGCGGTGCCAGCAGCCGAGCGGTGGGAAGCACCGAGACCGGCGG
29 34
61 GATGTCCCGCGCAGCAGCGCCAGCGGTGCACCCCGGGGACCGCGCGGACCGG
77 82
121 GTCGGAGTCGAC
...
-----
1 match found in sequence:
aaf25091 ; PCR primer for M. vaccae immunogenic epitope DNA9A.
(from "mycobacterng.seq")
TOIG of: aaf25091 check: 437 from: 1 to: 24
ID AAF25091 standard; DNA; 24 BP.
XX AC AAF25091;
XX DT 30-APR-2001 (first entry)
XX DE PCR primer for M. vaccae immunogenic epitope DNA9A.
XX KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX KW cancer; mycobacterial infection; TH1 immune response; vaccine;
XX KW PCR primer; ss.
XX OS Mycobacterium vaccae.
XX PN WO200104140-A1.
XX PD 18-JAN-2001.
XX DT 30-APR-2001 (first entry)
XX DE PCR primer for M. vaccae immunogenic epitope DNA9A.
XX KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX KW cancer; mycobacterial infection; TH1 immune response; vaccine;
XX KW PCR primer; ss.
XX OS Mycobacterium vaccae.
XX PN WO200104140-A1.
XX PD 18-JAN-2001.
XX DT 10-JUL-2000; 2000WO-NZ00121.
XX PF Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX PR vaccae, useful for treating mycobacterial infections, immune disorders
XX PR and cancers -
XX PR 12-JUL-1999; 99US-0351348.
XX PR 29-NOV-1999; 99US-0450072.
XX PR
```

```
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX PI Delcayre A;
XX XX WPI; 2001-168411/17.
XX DR Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX PT vaccae, useful for treating mycobacterial infections, immune disorders
XX PT and cancers -
XX PS Example 2; Page 61; 80pp; English.
XX XX The specification describes an immunogenic epitope of a Mycobacterium
XX CC vaccae antigen. The epitope is a stimulator of cytokine production.
XX CC The epitopes are useful for the treatment of immune disorders,
XX CC infectious diseases, especially tuberculosis, and cancer. They are
XX CC also useful for treatment of other mycobacterial infections such as
XX CC those caused by Mycobacterium avium. The epitopes are especially useful
XX CC for inducing TH1 immune responses, and for producing vaccines. PCR
XX CC primers AAF25090-91 were used to amplify DNA encoding a M. vaccae epitope
XX CC of the invention.
XX SQ Sequence 24 BP; 6 A; 9 C; 7 G; 2 T; 0 other;
AAF25091 Length: 24 March 5, 2002 14:18 Type: N Check: 437
Found using 'seq2-3' (pappu403.key)
1 TTACGCCCGAGAGAGCGCGAGGCC
13 18
-----
1 match found in sequence:
aaf25094 ; PCR primer used
(from "mycobacterng.seq")
TOIG of: aaf25094 check: 2709 from: 1 to: 34
ID AAF25094 standard; DNA; 34 BP.
XX AC AAF25094;
XX DT 30-APR-2001 (first entry)
XX DE PCR primer used to amplify hGH signal peptide-M. vaccae epitope DNA.
XX KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX KW cancer; mycobacterial infection; TH1 immune response; vaccine;
XX KW human growth hormone; hGH; PCR primer; ss.
XX OS Mycobacterium vaccae.
XX PN WO200104140-A1.
XX PD 18-JAN-2001.
XX DT 10-JUL-2000; 2000WO-NZ00121.
XX PF 12-JUL-1999; 99US-0351348.
XX PR 29-NOV-1999; 99US-0450072.
XX XX (GENE-) GENESIS RES & DEV CORP LTD.
XX PI Delcayre A;
XX XX WPI; 2001-168411/17.
XX DR Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX PT vaccae, useful for treating mycobacterial infections, immune disorders
XX PT and cancers -
XX XX Example 2; Page 62; 80pp; English.
XX PS
```

```
XX 18-JAN-2001.
XX 10-JUL-2000; 2000WO-NZ00121.
XX 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Delcayre A;
XX WPI; 2001-168411/17.
XX P-PSDB; AAB31654.
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX vaccae, useful for treating mycobacterial infections, immune disorders
XX and cancers -
XX Claim 4; Page 60; 80pp; English.
XX The specification describes an immunogenic epitope of a Mycobacterium
XX vaccae antigen. The epitope is a stimulator of cytokine production.
XX The epitopes are useful for the treatment of immune disorders,
XX infectious diseases, especially tuberculosis, and cancer. They are
XX also useful for treatment of other mycobacterial infections such as
XX those caused by Mycobacterium avium. The epitopes are especially useful
XX for inducing TH1 immune responses, and for producing vaccines. The
XX present sequence encodes a M. vaccae epitope of the invention.
XX Sequence 279 BP; 46 A; 98 C; 96 G; 39 T; 0 other;
SQ AAF25085 Length: 279 March 5, 2002 14:18 Type: N Check: 6231 ..
Found using 'seq2-3' (pappu403.key)
...
56 GGGTCCTCTGGCGCAGTACGCAAGTCGTGGCGCAGCATCGCGAACTGTGCTGCCG
106 111
116 GACAAAGTTCCTGGCGACACCGCGCAGCGCCCGGAGACCGCAGCGAGCGGATGTC
140 145
143 148
176 GCCGAACTGGATCCGTGGAGCGCGGGCTGAAGTCGGCCATCGCGCGGTGCGTGC
218 223
236 CGCGGGGCCCAAGTGTGCTGCGAAGGCGCGGAAGATCGCGAT
251 256
-----
5 matches found in sequence:
aaf25086; Nucleotide sequence of a M. vaccae immunogenic epitope from DNA44.
(from "mycobacteri.ng.seq")
TOIG of: aaf25086 check: 7203 from: 1 to: 132
ID AAF25086 standard; DNA; 132 BP.
XX AAF25086;
AC AAF25086;
XX 30-APR-2001 (first entry)
XX Nucleotide sequence of a M. vaccae immunogenic epitope from DNA44.
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX Mycobacterium vaccae.
OS
```

```
XX WO200104140-A1.
XX 18-JAN-2001.
XX 10-JUL-2000; 2000WO-NZ00121.
XX 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Delcayre A;
XX WPI; 2001-168411/17.
XX P-PSDB; AAB31655.
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
XX vaccae, useful for treating mycobacterial infections, immune disorders
XX and cancers -
XX Claim 4; Page 60; 80pp; English.
XX The specification describes an immunogenic epitope of a Mycobacterium
XX vaccae antigen. The epitope is a stimulator of cytokine production.
XX The epitopes are useful for the treatment of immune disorders,
XX infectious diseases, especially tuberculosis, and cancer. They are
XX also useful for treatment of other mycobacterial infections such as
XX those caused by Mycobacterium avium. The epitopes are especially useful
XX for inducing TH1 immune responses, and for producing vaccines. The
XX present sequence encodes a M. vaccae epitope of the invention.
XX Sequence 132 BP; 16 A; 44 C; 50 G; 22 T; 0 other;
SQ AAF25086 Length: 132 March 5, 2002 14:18 Type: N Check: 7203 ..
Found using 'seq2-3' (pappu403.key)
...
8 AACGGGCGGTCTGCGGGTTGAGGTCTCGTCCGCCAGTCGCCGTCCAGCACCGTGCCTCG
58 63 66
68 GCGCTGGTGATGCGCGCGCGGTAGCGGTCTCTCGTCCAGCACCGTGCCTCGCGCGG
72 91 96 107 112
128 CGGAT
-----
2 matches found in sequence:
aaf25087; Nucleotide sequence of a M. vaccae immunogenic epitope from DNA45.
(from "mycobacteri.ng.seq")
TOIG of: aaf25087 check: 4115 from: 1 to: 177
ID AAF25087 standard; DNA; 177 BP.
XX AAF25087;
AC AAF25087;
XX 30-APR-2001 (first entry)
XX Nucleotide sequence of a M. vaccae immunogenic epitope from DNA45.
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
XX cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX Mycobacterium vaccae.
XX WO200104140-A1.
XX 18-JAN-2001.
PD
```

```
DT 30-APR-2001 (first entry)
XX Nucleotide sequence of a M. vaccae immunogenic epitope from DNA29.
DE
XX
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX Mycobacterium vaccae.
OS
XX WO200104140-A1.
XX
XX 18-JAN-2001.
XX
XX 10-JUL-2000; 2000WO-NZ00121.
XX
XX 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Delcayre A;
XX
XX WPI; 2001-168411/17.
XX P-PSDB; AAB31652.
XX
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers
PT
XX
XX Claim 4; Page 60; 80pp; English.
XX
XX The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders,
CC infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes a M. vaccae epitope of the invention.
XX
XX Sequence 138 BP; 23 A; 44 C; 47 G; 24 T; 0 other;
XX
AAAF25083 Length: 138 March 5, 2002 14:18 Type: N Check: 6079
Found using 'seq2-3' (pappu403.key)
...
1 ATCGCGCGGCTGTGCGGGAGGAGGAGCGGTAGCGGGGTGCACACTACGTGCGCCCGGTT
|-----| |-----|
36 41 49 54
61 GCGGAGAGCAGGACTACATCGACCGAGCGCTTGCGCAACATCGG
...
-----
5 matches found in sequence:
aaf25084 ; Nucleotide sequence of a M. vaccae immunogenic epitope from DNA37.
(from "mycobactereng.seq")
TOIG of: aaf25084 check: 2916 from: 1 to: 261
ID AAF25084 standard; DNA; 261 BP.
XX
XX AAF25084;
XX
XX 30-APR-2001 (first entry)
XX
XX Nucleotide sequence of a M. vaccae immunogenic epitope from DNA37.
DE
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX Mycobacterium vaccae.
OS
XX WO200104140-A1.
XX
PN
```

```
XX 18-JAN-2001.
XX
XX 10-JUL-2000; 2000WO-NZ00121.
XX
XX 12-JUL-1999; 99US-0351348.
XX 29-NOV-1999; 99US-0450072.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Delcayre A;
XX
XX WPI; 2001-168411/17.
XX P-PSDB; AAB31653.
XX
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers
PT
XX
XX Claim 4; Page 60; 80pp; English.
XX
XX The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders,
CC infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes a M. vaccae epitope of the invention.
XX
XX Sequence 261 BP; 37 A; 94 C; 83 G; 47 T; 0 other;
XX
AAAF25084 Length: 261 March 5, 2002 14:18 Type: N Check: 2916
Found using 'seq2-3' (pappu403.key)
...
67 CGATGCGGAACACGACCGACCGAGCAGTGTCTCTCGGACCCCTGGGTACGC
|-----| |-----|
127 132 151 156
187 TGCGCGGTTGTTTCAGTCGTCGTCGCGGTGGCTGTTCTGCAACGCGCGCCCGCCCGT
|-----| |-----|
202 207 245
247 ---| CGTCCGTCGATACG
250
-----
5 matches found in sequence:
aaf25085 ; Nucleotide sequence of a M. vaccae immunogenic epitope from DNA42.
(from "mycobactereng.seq")
TOIG of: aaf25085 check: 6231 from: 1 to: 279
ID AAF25085 standard; DNA; 279 BP.
XX
XX AAF25085;
XX
XX 30-APR-2001 (first entry)
XX
XX Nucleotide sequence of a M. vaccae immunogenic epitope from DNA42.
DE
XX Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX Mycobacterium vaccae.
OS
XX WO200104140-A1.
XX
PN
```

TOIG of: aaf25081 check: 2826 from: 1 to: 291

ID AAF25081 standard; DNA; 291 BP.
XX AAF25081;
AC AAF25081;
XX AAF25081;
DT 30-APR-2001 (first entry)
XX AAF25081;
DE Nucleotide sequence of a M. vaccae immunogenic epitope from DNA26.
XX AAF25081;
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX AAF25081;
OS Mycobacterium vaccae.
XX AAF25081;
PN WO200104140-A1.
XX AAF25081;
PD 18-JAN-2001.
XX AAF25081;
PF 10-JUL-2000; 2000WO-NZ00121.
XX AAF25081;
PR 12-JUL-1999; 99US-0351348.
XX AAF25081;
PR 29-NOV-1999; 99US-0450072.
XX AAF25081;
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX AAF25081;
PI Delcayre A;
XX AAF25081;
DR WPI; 2001-168411/17.
XX AAF25081;
DR P-PSDB; AAB31651.
XX AAF25081;
PT Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -
XX AAF25081;
PS Claim 4; Page 59; 80pp; English.
XX AAF25081;
CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes a M. vaccae epitope of the invention.
XX AAF25081;
SQ Sequence 291 BP; 45 A; 103 C; 101 G; 42 T; 0 other;

AAF25081 Length: 291 March 5, 2002 14:18 Type: N Check: 2826 ..
Found using 'seq2-3' (pappu403.key)

1 ATCAGTTCGGCCCTGTCGCCACGCCGCGAGGCGAGTCCTCCGCTCCGGCGTCGATC
16 21 |-----| |-----|
51 56
61 GGGTTGGTCCGTCGGCCGACACACACATCCACCCGAGGTCGAGCAACGGGTCCTCCG
121 ACGGTGCACATCTCCAGTCGATGACGCCGCGAGCTCGGGGACGTCGGGGCGGACGAGC
145 150 |-----| |-----|
162 167
181 ACGTTGTTTCAGATGGCAGTCGCGCGTCGATGATCCCGGGGTCGCGGCTCGGCGCTGCGC
198 203 |-----| |-----|
222 227
241 GAGTCAGCCAGTCGGCGGAGACATGCACCCGAGGGAACGACTCGGGGCGCG
277 282 |-----|
1 match found in sequence:

aaf25082 ; Nucleotide sequence of a M. vaccae immunogenic epitope from DNA27.
(from "mycobactering.seq")
TOIG of: aaf25082 check: 2042 from: 1 to: 147

ID AAF25082 standard; DNA; 147 BP.
XX AAF25082;
AC AAF25082;
XX AAF25082;
DT 30-APR-2001 (first entry)
XX AAF25082;
DE Nucleotide sequence of a M. vaccae immunogenic epitope from DNA27.
XX AAF25082;
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX AAF25082;
OS Mycobacterium vaccae.
XX AAF25082;
PN WO200104140-A1.
XX AAF25082;
PD 18-JAN-2001.
XX AAF25082;
PF 10-JUL-2000; 2000WO-NZ00121.
XX AAF25082;
PR 12-JUL-1999; 99US-0351348.
XX AAF25082;
PR 29-NOV-1999; 99US-0450072.
XX AAF25082;
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX AAF25082;
PI Delcayre A;
XX AAF25082;
DR WPI; 2001-168411/17.
XX AAF25082;
DR P-PSDB; AAB31651.
XX AAF25082;
PT Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -
XX AAF25082;
PS Claim 4; Page 59; 80pp; English.
XX AAF25082;
CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes a M. vaccae epitope of the invention.
XX AAF25082;
SQ Sequence 147 BP; 25 A; 54 C; 49 G; 19 T; 0 other;

AAF25082 Length: 147 March 5, 2002 14:18 Type: N Check: 2042 ..
Found using 'seq2-3' (pappu403.key)

3 CACGAGGTAGGCGCGTCCAGCCGCTACTCTTCGCCCCAGACAGCGGTGCGCGCGCGC
53 58 |-----|
63 AGACCAGCGTCTCTCGCGCCAGATACACCCAGCGGTGGCGCGCAT
...
2 matches found in sequence:
aaf25083 ; Nucleotide sequence of a M. vaccae immunogenic epitope from DNA29.
(from "mycobactering.seq")
TOIG of: aaf25083 check: 6079 from: 1 to: 138
ID AAF25083 standard; DNA; 138 BP.
XX AAF25083;
AC AAF25083;
XX AAF25083;

2 matches found in sequence:
aaf25071 ; Nucleotide sequence of a M. vaccae immunogenic epitope.
(from "mycobacterng.seq")
TOIG of: aaf25071 check: 706 from: 1 to: 146

ID AAF25071 standard; DNA; 146 BP.
XX
AC AAF25071;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a M. vaccae immunogenic epitope.
XX
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX
OS Mycobacterium vaccae.
XX
FH Key Location/Qualifiers
FT CDS 7..144
FT /*tag= a
FT /note= "no termination codon given"
XX
PN WO200104140-A1.
XX
PD 18-JAN-2001.
XX
PF 10-JUL-2000; 2000WO-NZ00121.
XX
PR 12-JUL-1999; 99US-0351348.
PR 29-NOV-1999; 99US-0450072.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Delcayre A;
XX
DR WPI; 2001-168411/17.
DR P-PSDB; AAB31640.
XX
XX Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -
XX
PS Example 1; Page 58; 80pp; English.
XX
CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes a M. vaccae epitope of the invention.
XX
SQ Sequence 146 BP; 21 A; 49 C; 36 G; 40 T; 0 other;
AAAF25071 Length: 146 March 5, 2002 14:18 Type: N Check: 706
Found using 'seq2-3' (pappu403.key)

1 AAGCTTATGGCTACAGGCTCCCGAGCTCCCTGCTCTCTGGCTTTTGGCTGCTCTGCCCTG
24 29
|-----|
118

61 CCTGTGCTTCAAGAGGCGAGTGCCTTCCCAACCATTCCTTATCCAGGCTTTTGGACAAC
|---|
118

121 GCTATGACGCTGTGGCTTCGGGATCC
123
|-----|
2 matches found in sequence:
aaf25073 ; Nucleotide sequence of a M. vaccae immunogenic epitope.
(from "mycobacterng.seq")

aaf25072 ; Nucleotide sequence of a M. vaccae immunogenic epitope.
(from "mycobacterng.seq")
TOIG of: aaf25072 check: 8435 from: 1 to: 145

ID AAF25072 standard; DNA; 145 BP.
XX
AC AAF25072;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a M. vaccae immunogenic epitope.
XX
KW Epitope; antigen; cytokine production; immune disorder; tuberculosis;
KW cancer; mycobacterial infection; TH1 immune response; vaccine; ss.
XX
OS Mycobacterium vaccae.
XX
FH Key Location/Qualifiers
FT CDS 7..144
FT /*tag= a
FT /note= "no termination codon given"
XX
PN WO200104140-A1.
XX
PD 18-JAN-2001.
XX
PF 10-JUL-2000; 2000WO-NZ00121.
XX
PR 12-JUL-1999; 99US-0351348.
PR 29-NOV-1999; 99US-0450072.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Delcayre A;
XX
DR WPI; 2001-168411/17.
XX
PT Novel polypeptides comprising immunogenic epitopes of Mycobacterium
PT vaccae, useful for treating mycobacterial infections, immune disorders
PT and cancers -
XX
PS Example 1; Page 58; 80pp; English.
XX
CC The specification describes an immunogenic epitope of a Mycobacterium
CC vaccae antigen. The epitope is a stimulator of cytokine production.
CC The epitopes are useful for the treatment of immune disorders.
CC Infectious diseases, especially tuberculosis, and cancer. They are
CC also useful for treatment of other mycobacterial infections such as
CC those caused by Mycobacterium avium. The epitopes are especially useful
CC for inducing TH1 immune responses, and for producing vaccines. The
CC present sequence encodes a M. vaccae epitope of the invention.
XX
SQ Sequence 145 BP; 21 A; 49 C; 35 G; 40 T; 0 other;
AAAF25072 Length: 145 March 5, 2002 14:18 Type: N Check: 8435
Found using 'seq2-3' (pappu403.key)

1 AAGCTTATGGCTACAGGCTCCCGAGCTCCCTGCTCTCTGGCTTTTGGCTGCTCTGCCCTG
24 29
|-----|
118

61 CCTGTGCTTCAAGAGGCGAGTGCCTTCCCAACCATTCCTTATCCAGGCTTTTGGACAAC
|---|
118

121 GCTATGACGCTGTGGCTTCGGATCC
123
|-----|
2 matches found in sequence:
aaf25073 ; Nucleotide sequence of a M. vaccae immunogenic epitope.
(from "mycobacterng.seq")

XX (SUMU) SUMITOMO PHARM CO LTD.
XX Scarlato GR, Hadida Ruah SS, Nishimura T, Nakatsuka M, Samizo F;
PI Kamikawa Y, Houtigai H;
XX WPI: 2000-656423/63.
XX New hydroxamic acid derivatives are useful in the treatment of e.g.
PT acne, anorexia, cardiac infarction and gum disease and are matrix
PT metallo-proteinase inhibitors -
XX Disclosure; Page 200; 218pp; English.
XX The present sequence is a PCR primer used to sublone the C-terminal
CC truncated form of matrix metalloproteinase-13. This was used for
CC inhibition assays in an example of an invention relating to hydroxamic
CC acid derivatives and their salts. These are used for treating diseases
CC associated with excess or undesired matrix metalloproteinases,
CC especially matrix metalloproteinase 3 and/or 13. They may be used to
CC treat abnormal wound healing, acne, acute coronary syndrome, acute
CC infection, AIDS, alcoholism, allergic conjunctivitis, allergic reactions,
CC allergic rhinitis, ALS, Alzheimer's disease, anaphylaxis, aneurysmal
CC aortic disease, angina, angiofibromas, anorexia, aortic aneurysm, ARDS,
CC aspirin-independent anti-thrombosis, asthma, atherosclerosis,
CC atherosclerotic plaque rupture, atopic dermatitis, benign hyperplasia,
CC bleeding, bone fractures, bronchitis, burns, cachexia, cancer, cardiac
CC infarction, cardiac insufficiency, cardiomyopathy, cerebral
CC haemorrhaging, cerebral ischaemia, cerebral vascular dementia, CHF,
CC chronic bronchitis, chronic dermal wounds, chronic obstructive
CC pulmonary disease, cirrhosis, congestive heart failure, corneal injury,
CC coronary thrombosis, Crohn's disease, cystic fibrosis, decubitus ulcer,
CC diabetic disorders, Duchenne's muscular dystrophy, emphysema,
CC endometriosis, endosclerolysis, epidermolysis bullosa, eye disorders,
CC fibrosis, gastritis, gingivitis, glomerular diseases,
CC glomerulonephritis, gout, graft rejection, gum disease, GVHD,
CC Hashimoto's thyroiditis, head trauma, headaches, heart attacks, heart
CC failure, haemangiomas, haemorrhage, hepatitis, hirsutism, Huntington's
CC disease, hypertension, insulin resistance, interstitial nephritis,
CC ischaemia, ischaemic heart disease, Kaposi's sarcoma, keratinisation,
CC keratitis, kidney failure, leishmaniasis, leprosy, leukaemia, leukocyte
CC infiltration, liver cirrhosis, loss of appetite, macular degeneration,
CC malaria, mandibular joint disease, memory impairment, meningitis,
CC migraine, miscarriage, multi-infarct dementia, multiple sclerosis,
CC muscular dystrophy, myalgia, myasthenia gravis, myelinic degradation,
CC myocardial infarction, myopia, neovascular glaucoma, neuroinflammation,
CC ocular tumours, optic neuritis, osteoarthritis, osteopaenia, Paget's
CC disease, pain, pancreatitis, Parkinson's disease, periodontitis,
CC peripheral vascular disease, polyarteritis nodositas, polychondritis,
CC premature childbirth, premature rupture of foetal membranes, prion
CC disease, proliferative retinopathies, proteinurea, pseudogout,
CC psoriasis, pterygium, pulmonary emphysema, radiation damage, reperfusion
CC snake bite, Reiter's syndrome, renal fibrosis, reocclusion, senility,
CC injury, restenosis, scleritis, scleroderma, senile dementia, senility,
CC sepsis, septic shock, Sharp syndrome, Sjogren's syndrome, SLE,
CC spondylitis, stenosis, sterility, stroke, system sclerosis, thrombosis,
CC toxic effects of chemotherapy, toxic shock, tuberculosis, ulcerations,
CC ulcerative colitis, uraemia, vasculitis, ventricular dilation, vesicular
CC epidermolysis.
XX
SQ Sequence 51 BP; 11 A; 12 C; 11 G; 17 T; 0 other;
AAC65329 Length: 51 March 5, 2002 14:18 Type: N Check: 6058 ..
Found using 'seq2-3' (pappu403.key)

1 TTTGGATCCTTAGCCGACAGCGCTTTGAATACCTTGATCATCGTCATCAGG
40 45
|-----|

1 match found in sequence:
aaf24075 : Wnt-3a gene specific primer wn3a.2.
(from "mycobacterng.seq")

TOTG of: aaf24075 check: 4831 from: 1 to: 20
ID AAF24075 standard; DNA; 20 BP.
AC AAF24075;
XX
DT 28-MAR-2001 (first entry)
XX
XX Wnt-3a gene specific primer wn3a.2.
DE
DE
KW Mouse; Wnt; antianaemic; cytostatic; haemostatic; myelopoiesis;
KW mythropoiesis; lymphopoiesis; lymphocytopaenia; lymphostasis;
KW immunodeficiency; infection; erythrocytopaenia; erythroblastopaenia;
KW leukoerythroblastosis; erythroclasis; thalassemia; anaemia;
KW disseminated intravascular coagulation; DIC; myelodysplasia; haemorrhage;
KW thrombocytosis; immune thrombocytopaenia purpura; ITP; PCR primer; ss.
XX
OS Mus sp.
XX
PN US6159462-A.
XX
XX 12-DEC-2000.
XX
XX 15-AUG-1997; 97US-0911860.
XX
XX 16-AUG-1996; 96US-0024068.
XX (GETH) GENENTECH INC.
XX Matthews W, Austin TW;
XX
XX WPI: 2001-070102/08.
XX
XX Use of Wnt polypeptides for enhancing or increasing myelopoiesis,
PT erythropoiesis or lymphopoiesis, for treating lymphopaenia,
PT lymphorrhea, anemia, autoimmunethrombocytopenio purpura -
XX
XX Example 1: Column 48; 40pp; English.
XX
XX The present sequence is a primer used in the analysis of Wnt genes in
CC cells and cell lines. This was performed as an example in a
CC specification relating to the use of Wnt polypeptides for enhancing or
CC increasing myelopoiesis, erythropoiesis or lymphopoiesis. The method
CC comprises exposing haematopoietic stem/progenitor cells to Wnt
CC polypeptides for stimulating proliferation or differentiation of the
CC cells. Enhancement of lymphopoiesis may be useful for improving the
CC disease conditions of lymphocytopaenia, lymphorrhea, lymphostasis,
CC immunodeficiency (e.g. HIV and AIDS), infections (including, for example,
CC opportunistic infections and tuberculosis (TB)), lupus, and other
CC disorders characterised by lymphocyte deficiency. Enhancement of
CC erythropoiesis may be beneficial for the disease conditions
CC erythrocytopaenia, erythrodegenerative disorders, erythroblastopaenia,
CC leukoerythroblastosis, erythroclasis, thalassemia and anaemia (e.g.
CC haemolytic anaemia, such as acquired, autoimmune or
CC microangiopathicchemolytic anaemia, aplastic anaemia, congenital anaemia,
CC refractory anaemia, radiation anaemia, sickle cell anaemia, splenic
CC anaemia and toxic anaemia etc. Enhancement of myelopoiesis is beneficial
CC to disseminated intravascular coagulation (DIC), immune (autoimmune)
CC thrombocytopaenia purpura (ITP); HIV induced ITP, myelodysplasia,
CC thrombocytotic diseases and thrombocytosis. The Wnt polypeptides are
CC also useful in improving engraftment in bone marrow transplantation
CC and in treating patients having suffered a haemorrhage.
XX
SQ Sequence 20 BP; 2 A; 10 C; 6 G; 2 T; 0 other;
AAF24075 Length: 20 March 5, 2002 14:18 Type: N Check: 4831 ..
Found using 'seq2-3' (pappu403.key)

1 CAGCCCGCGCGTCGCTC
8 13
|-----|

XX AAA98933;
AC
XX
DT 16-FEB-2001 (first entry)
XX
DE
DE H. tuberculata hemocyanin HcH2 domain h'' encoding DNA.
XX
XX Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
XX Hallotis tuberculata.
OS
XX
XX WO200055192-A2.
PN
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
PI
XX
XX WPI; 2000-587517/55.
DR
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
XX Claim 1a; Page 157; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
XX Sequence 1548 BP; 438 A; 401 C; 331 G; 377 T; 1 other;
SQ
AAA98933 Length: 1548 March 5, 2002 14:18 Type: N Check: 2031
Found using 'seq2-3' (pappu403.key)
...
92 TGTCCTGGATGAAGCAACAACTTGAGAATGCCCTTTACAAGCTACAGACGACACA
142 147
152 GTCTAACGGGATACGAAGCAATCTCTGGTTACCATTGATATACCCCAATCTGTGTCGCGAAG
212 AAGGCGATGACAAAAATACCCCTGCTGCGTCCCGGATGGGCATCTTTCTCTTACTGGCA
214 219
...

307 CACAATGGTGCACACTGCTTGGTGTCTTACTGGGACTGGAACAAGGACCTGCTGCTCACTG
357 362
367 CCGCGCTTCTTCTCCGACTCCAGCAACAACAACTTCAAGTACCAACGCGCGGT
369 374
427 GTTGGTCACGACACACCGCTCAGAGAGCCAACTAGTCTTATATATAACCA
...
524 TTGAAGAAACAATTACTGGGACTTTGAGGTTTCAGTATGAGATCCTCCACAAACGCGTCC
574 579
584 ACTCCTGGCTTGGAGGATCCAGAAAGTATTCATGTCTACCCCTGGA
...
721 AAGCCCTACAACCTTCGCTAAATGTGCTTATCATATGATGAAGACCACTGGCGCCCTTC
771 776
781 AGCTATCCATCTATCAACACGAGGAGGTTCACCGCTGCCAACTCCA
...
1256 TGTTCTTGTATTCCACGATANGAAGAAACCTTCACATCCCTCCCAAGTTGTCTGTCGAAGA
1306 1311
1316 AAGGCAACCCGATCGAGTTCCACCCAGTCGATGATTCAGTTACGAG
...

1 match found in sequence:
aac65329 ; MMP-13 cDNA oligonucleotide primer #2.
(from "mycobacterng.seq")
TOIG of: aac65329 check: 6058 from: 1 to: 51
ID AAC65329 standard; DNA; 51 BP.
XX
AC AAC65329;
XX
DT 09-FEB-2001 (first entry)
XX
DE MMP-13 cDNA oligonucleotide primer #2.
XX
KW Human; matrix metalloproteinase-13; MMP-13; procollagenase-3;
KW vulvary; antisecretory; dermatological; antibacterial; anti-HIV;
KW antialcoholic; antiallergic; antiinflammatory; nootropic;
KW neuroprotective; immunosuppressive; antiangiogenic; anticoagulant;
KW thrombolytic; antiasthmatic; antiarteriosclerotic; cytostatic;
KW immunomodulator; cardiant; hepatotropic; antiulcer; antidiabetic;
KW ophthalmological; gynecological; nephrotropic; antigout;
KW immunosuppressive; thyromimetic; haemostatic; hepatotropic; virucide;
KW depilatory; anticonvulsant; hypotensive; vasotrophic; analgesic;
KW antiemigraine; osteopathic; antiarthritic; antipruritic; antitumor;
KW antiparkinsonian; tocolytic; antipsoriatic; antinfertility;
KW cerebroprotective; tuberculostatic; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200063197-A1.
XX
PD 26-OCT-2000.
XX
PF 19-APR-2000; 2000WO-US10383.
XX
PR 19-APR-1999; 99US-0129933.

```
2      GTGACATAAATACAGGAGCATGTCACCGAACCGGTGTTCCGGTGAGCGCATCGT      |-----|
      52  57
62      CTGGGAGGACCTGTCTAGTCTCAAGTCTGCTCTGCGAGACCTACA
      ...
239     .GACTGTACACCTCCAGTTGGAGATGGCTCTGAGGAGACATGGATCATCTCTCGCCATCC      |-----|
      289 294
299     CCTACTGGGACTGGACAAAGCCTATCTCCGAACCTCCCTCGCTCTT
      ...
540     TGAGGTACAGTTTGAGGTCTCCATACAGTGATCCACTACCTTGTGTGGAGCTGCAGAC      |-----|
      590 595
600     CTACGCATTGCTTCTCTGCATTATGCTCTCTACGACCCATTCTTC
      ...
1153    CCATTTTATCAAGTTGAGATCCATGCTGTTAACAGACCATGATACCGTCTGCTCTGTG      |-----|
      1203 1208
1213    ATCCAGCCCCCACTATCATCTATTCTCTCTGGGAA
-----
6 matches found in sequence:
aaa98932 ; H. tuberculata hemocyanin Hth2 domain g'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98932 check: 3563 from: 1 to: 1206

ID      AAA98932 standard; DNA; 1206 BP.
XX
AC      AAA98932;
XX
DT      16-FEB-2001 (first entry)
XX
DE      H. tuberculata hemocyanin Hth2 domain g'' encoding DNA.
XX
KW      Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW      immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW      infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW      breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW      pharmaceutical carrier; ds.
XX
OS      Haliotis tuberculata.
XX
PN      WO200055192-A2.
XX
PP      21-SEP-2000.
XX
PF      17-MAR-2000; 2000WO-EP02410.
XX
PR      17-MAR-1999; 99DE-1011971.
XX
PS      20-AUG-1999; 99DE-1039578.
XX
PA      (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI      Markl J, Altenhein B, Lieb B, Stiefel T;
XX
DR      WPI; 2000-587517/55.
XX
PT      New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX      and for recombinant production of fusion proteins for vaccination
XX
PS      Claim 1a; Page 156-157; 163pp; German.
XX
CC      This invention describes a novel nucleic acid (I) containing a sequence
```

```
CC      that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC      immunological properties of at least one domain of (II). The products of
CC      the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC      immunomodulatory and antihypertensive activity (I), and constructs
CC      additionally containing antigen-encoding sequences, are useful in gene
CC      therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC      parasitic or viral infections and tumors, particularly schistosomiasis
CC      and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC      colon-rectum), also hypertension, as vaccines, for treating cocaine
CC      misuse and very generally as carriers for pharmaceuticals, e.g.
CC      cytostatics. They may also be used to generate antibodies (Ab). Probes
CC      based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC      (by detecting specific binding to cellular DNA or proteins), particularly
CC      where associated with the types of carcinoma listed above. Hemocyanins
CC      can be produced recombinantly, relatively inexpensively and in adequate
CC      amounts, eliminating the need to culture gastropods. When used as a
CC      carrier, (II) significantly increases the half-life of the attached
CC      pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ      Sequence 1206 BP; 302 A; 335 C; 282 G; 287 T; 0 other;
AAA98932 Length: 1206 March 5, 2002 14:18 Type: N Check: 3563
Found using 'seq2-3' (pappu403.key)
...
8      CTGCTGACAGTGGCGACTCTGCCAACATTGCTGCTCTGGGTGAGGAAGACGCTCACGA      |-----|
      58  63
68      CCTCACTGTGCTGAGACCGAGAACCTAAGACAGAGGCTCTTCAAGG
      ...
392     TTGATCATCTCGGTGTAACACAGTCAAGTTCCTCCCGCAGAGACATGCTGTTTAAACGCCAG      |-----|
      442 447
452     AGCAAGGATCAGAGTCTTCTTCTATAGACAAGTCTCTCTGGCTTTGGAGCAGACTGACT      |-----|
      465 470
512     ACTCCAGTTCGAAGTCCAGTTTGAGTGAAGCCACAGCCCATTCACCTCTGGACAGGTG      |-----|
      547 552
572     GACGTAGCCCTTACGAATGTCGACCCCTCGA
      ...
704     ACGAAGCACACTGTGAATCCAGGTTCCTGAACACAGCCCTTGAGGCCATTCAACGATGACA      |-----|
      754 759
764     TCAACCACAATCCCAATCACCAGAACTAATGCCAGGCCCTATCGATTTC
      ...
911     AGAGAACATTGTGTCCTCTCTTCTCGTGAATCGGTGCAGTGTGATGTCGCTCTTTG      |-----|
      961 966
971     ACATCTCGCGCCCAATGGTGACTGTGTCTTTTTCGAGGAACCTTTTC
      ...
-----
8 matches found in sequence:
aaa98933 ; H. tuberculata hemocyanin Hth2 domain h'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98933 check: 2031 from: 1 to: 1548

ID      AAA98933 standard; DNA; 1548 BP.
```

DE H. tuberculata hemocyanin Hth2 domain e'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
XX WO200055192-A2.
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
XX
XX 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination -
XX
XX Claim la; Page 155-156; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
XX that encodes hemocyanin (II), a domain of (I) or its fragment with the
XX immunological properties of at least one domain of (II). The products of
XX the invention have cytostatic, virucide, antibacterial, antiparasitic,
XX immunomodulatory and antihypertensive activity. (I), and constructs
XX additionally containing antigen-encoding sequences, are useful in gene
XX therapy of tumors. Polypeptides encoded by (I) are useful for treating
XX parasitic or viral infections and tumors, particularly schistosomiasis
XX and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
XX colon-rectum), also hypertension, as vaccines, for treating cocaine
XX misuse and very generally as carriers for pharmaceuticals, e.g.
XX cytostatics. They may also be used to generate antibodies (Ab). Probes
XX based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
XX (by detecting specific binding to cellular DNA or proteins), particularly
XX where associated with the types of carcinoma listed above. Hemocyanins
XX can be produced recombinantly, relatively inexpensively and in adequate
XX amounts, eliminating the need to culture gastropods. When used as a
XX carrier, (II) significantly increases the half-life of the attached
XX pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1257 BP; 332 A; 304 C; 290 G; 331 T; 0 other;

AAA98930 Length: 1257 March 5, 2002 14:18 Type: N Check: 1250
Found using 'seq2-3' (pappu403.key)

482 TCAAAGTGGACAGCTTTTGTATAAAGTGTCTTCAACAAGCCATCTTCAGCTCAGC
|-----|
532 537

542 AGGAAACTACTGTGACTTTTCAGATTTCAGTTTGAAATCTTCACAACGGCGTTCACACGT
|-----|
589 594

602 GGSTCGGAGCAGTCGTACCTACTCTATCGGACATCTTCATTA

749 CTCACCTGTCTCTCGAGAAATGAGAACCAATTGAAGCCTTTCAGCTTCGGCGCTCCTT
|-----|
799 804

809 ATAACTGGATCAGCTCACAGGATTTTCTCCCGACCCGAGGACAC

...

4 matches found in sequence:
aaa98931 ; H. tuberculata hemocyanin Hth2 domain f'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98931 check: 5205 from: 1 to: 1248

ID AAA98931 standard; DNA; 1248 BP.
XX
XX AC AAA98931;
XX
XX 16-FEB-2001 (first entry)
XX
XX H. tuberculata hemocyanin Hth2 domain f'' encoding DNA.

DE Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
XX Haliotis tuberculata.

OS

XX WO200055192-A2.
XX
XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-EP02410.

XX 17-MAR-1999; 99DE-1011971.

XX 20-AUG-1999; 99DE-1039578.

XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.

XX Markl J, Altenhein B, Lieb B, Stiefel T;

XX WPI; 2000-587517/55.

XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination -
XX
XX Claim la; Page 156; 163pp; German.

XX This invention describes a novel nucleic acid (I) containing a sequence
XX that encodes hemocyanin (II), a domain of (I) or its fragment with the
XX immunological properties of at least one domain of (II). The products of
XX the invention have cytostatic, virucide, antibacterial, antiparasitic,
XX immunomodulatory and antihypertensive activity. (I), and constructs
XX additionally containing antigen-encoding sequences, are useful in gene
XX therapy of tumors. Polypeptides encoded by (I) are useful for treating
XX parasitic or viral infections and tumors, particularly schistosomiasis
XX and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
XX colon-rectum), also hypertension, as vaccines, for treating cocaine
XX misuse and very generally as carriers for pharmaceuticals, e.g.
XX cytostatics. They may also be used to generate antibodies (Ab). Probes
XX based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
XX (by detecting specific binding to cellular DNA or proteins), particularly
XX where associated with the types of carcinoma listed above. Hemocyanins
XX can be produced recombinantly, relatively inexpensively and in adequate
XX amounts, eliminating the need to culture gastropods. When used as a
XX carrier, (II) significantly increases the half-life of the attached
XX pharmaceutical, by inhibiting ultrafiltration in the kidneys.

SQ Sequence 1248 BP; 326 A; 313 C; 290 G; 319 T; 0 other;

AAA98931 Length: 1248 March 5, 2002 14:18 Type: N Check: 5205
Found using 'seq2-3' (pappu403.key)

...

415 GTTGAAGTTGCCCTTTGAATAGGAGCTACAGAACCTCACCAGATAGATGTTTGA
446 451
475 CAACCTTTATTTGGCAACATACGCGT
...
652 ATCTTCTACCTTCGCTCACTCAACACTGACCGGCTCTGGCAATTTGGCAAGCGTTGCAG
702 707
712 ATACGAACAAAGCCCTTACAAGCTCATTTGCTGCTGAGGAACGCCAGCTCTC
758 763
772 AAACCTTTCGCTTCAGTTCCCACTGAACAACAAACAACTACGAAACTCGGTG
832 CCCAACAGCTTTAGGACTACGAGGAGTCTCTGGCTATACTTATGATGACCTCAACTTC
838 843
892 GG
...
5 matches found in sequence:
aaa98929 ; H. tuberculata hemocyanin Hth2 domain d'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98929 check: 19 from: 1 to: 1242
ID AAA98929 standard; DNA; 1242 BP.
XX
AC
XX
XX
DT
XX
XX
DE
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination -
XX
XX Claim 1a; Page 155; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,

CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1242 BP; 338 A; 296 C; 276 G; 332 T; 0 other;
AAA98929 Length: 1242 March 5, 2002 14:18 Type: N Check: 19 ..
Found using 'seq2-3' (pappu403.key)
...
74 TGTCAAGGGAGAGTAGAGAGCCCTAAGGTCTGCCTTCTGCAACTCAGAACGACGAG
124 129
134 TCTATGAGATATTGCCAAGTTCCACAGGCAAGCCTGGGTGTGTGA
...
331 GAGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCAATTCCCGTCAACAACGTTTCAGC
381 386
391 CCTAATCCTTCTTCAGAGGTAAATCAGTTTGTGAGATGCTGTTA
437 CAACACGCTGATCCCGCAGCCTGAGCTGTACGTTAACAGGTACTACTACCAACGTCATGT
487 492
497 TGGTTTTTGACAGGACAACTACTCGGACTTCGAGATACAGTTTGA
...
864 GTTCAATCATTTTCAGTATTCGCCAGGCTTGAAGAAATCATTCGTATTAGACACGTCRAGA
914 919
924 TCGTGTGTTTGCAGGATTCCTCTTCACAACATTGGGACATCCCGA
...
1018 TGTGAAACAACAGCCGGAACATTTGCCGTACTCGGAGGAGAAACAGAGATGGCGTTTCAT
1068 1073
1078 TTTGACAGACTCTACAGGTTTGACATCAGTCAGTGAACACTGAGGGACC
...

3 matches found in sequence:
aaa98930 ; H. tuberculata hemocyanin Hth2 domain e'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98930 check: 1250 from: 1 to: 1257
ID AAA98930 standard; DNA; 1257 BP.
XX
XX AAA98930;
XX
XX 16-FEB-2001 (first entry)
XX

PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI: 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination -
XX
PS Claim 1a; Page 154; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 591 BP; 145 A; 181 C; 134 G; 131 T; 0 other;

AAA98926 Length: 591 March 5, 2002 14:18 Type: N Check: 5557 ..
Found using 'seq2-3' (pappu403.key)

...

107 GGTATCGTGGAAACATCAAGTTTGAGAAATGAAGACTGCAAGAGCTGTTGACGATCGCC
157 162 166
|-----|
|-----|

167 TTTTCGAGAAGGTTGGACCAGGAGAGAAATACCCGACTCTTTGAAGGAATT

...

227 TTGAACAGGATGAATTCTGCAACTTCGAGATCCAGTTTGAGTTGGCTCAACAGCTATCC
277 282
|-----|

287 ACTACCTGTTGGGCGCGTCACACGTACTCCATGTCTCATCTCGA

...

351 CCTCTTCTCTCCTCATCACTCCAACCGGACCGCATCTTCGCCATCTGGGAAGCTCTTCA
401 406
|-----|

411 GGTAAGGAGAAAGGACCCCAACACCCGCGGACTGCGCACACAAAC

...

7 matches found in sequence:
aaa98927 ; H. tuberculata hemocyanin Hth2 domain b'' encoding DNA.
(from "mycobacterng.seq")
ToIG of: aaa98927 check: 2766 from: 1 to: 1245

ID AAA98927 standard; DNA; 1245 BP.

XX

AC AAA98927;

XX

DT 16-FEB-2001 (first entry)

XX H. tuberculata hemocyanin Hth2 domain b'' encoding DNA.
DE
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
XX Hallotis tuberculata.
OS
XX
XX WO200055192-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 17-MAR-2000; 2000WO-EP02410.
PF
XX
XX 17-MAR-1999; 99DE-1011971.
PR
XX 20-AUG-1999; 99DE-1039578.
PR
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
PA
XX Markl J, Altenhein B, Lieb B, Stiefel T;
PI
XX
XX WPI: 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination -
XX
XX Claim 1a; Page 154; 163pp; German.
PS
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1245 BP; 314 A; 323 C; 299 G; 309 T; 0 other;

AAA98927 Length: 1245 March 5, 2002 14:18 Type: N Check: 2766 ..
Found using 'seq2-3' (pappu403.key)

...

175 TGTCCATTCCTCCGAGGCCAACAAATAGTTTCGCCTGTTCATCCACGCATGGCGCATTC
225 230
|-----|

235 CCTCATTTGGCAGACACTGTTTCGTACCCAGGTGGAAGATGCTCTGATCAGCGGAGGATCG
291
|-----|

295 -| CCTATAGGGGTCCCTACTGGGACTGGACTCAGCCTATGGCGCATCTCCAGGACTTGCA
296

355 GACACGCCACCTTATAGATCCCATCAGCGGGGACAGACACAAACCCCTTCCACGAT
358 363
|-----|

DE H. tuberculata hemocyanin Hth1 domain h'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
DR
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
PS Claim 1a; Page 153-154; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1536 BP; 440 A; 366 C; 357 G; 373 T; 0 other;
AAA98925 Length: 1536 March 5, 2002 14:18 Type: N Check: 2420 ..
Found using 'seq2-3' (pappu403.key)
...
86 CCCTGCAAGAGGCCAACCAATTAAGGATGCACGTACAGCTCCAGATCAGCAGCAGTA
136 141
146 AAGGGGGCTTTGAGGCCATAGCTGGCTATCAGGGGTATCCTAAATAT
...
515 TGGAGGAAAACCTCGTACTGTGACTTTGAAGTTCAGTATGAGATCCTCCATACGCCGTCC
565 570
575 ACTCCTGGCTTGAGGAACTGGAAGATATCCATGTCTACCCCTGGA
...

689 AGAAGTTGCAAAAGATAAGAATGAAGCCTTACTACGCATTGGATTGTGCTGGCAGAC
739 744
749 TTATGAAAGACCCCTGTCATCCCTTCAACTACGAAACCGTTTAATGA
...
959 TGTGGGGCTTACGGATATCAGCTACAGTGAAAGTATTATTTCATTTCGAAAAACGATACAA
1009 1014
1019 GTCAGGAAGAATATGCAGGAGAAATTGCAGTTTTTGGGAGGTGAGAA
...
1133 TGAAGATGAAGACATCCGTTTTTAGAGTGGTTGTACTGCCTACAAACGGTTCAGCTGTGTA
1183 1188
1193 CCACGAGGCTGCTCTCAGCCATTTCATCGTCCACCGTCCAGCCCATGTGGCTCAGCATCT
1216 1221
1253 TGGTAATCCCAAGTAGGTGC
...
1346 TGAACAAGCAATGCTGGAGCTGGCAGCTATAGTCTATGGTAAATGCATCGTTCGCC
1396 1401
1406 CTTTCTCTTACCACGGCTTTGAACCTGGGACAAAGTCTACAGCGTCGATCAGCGAGACTACT
1444 1449
1466 ACATTGCTGCAGGTACCCACGCGTTGTGTGAGCA
...

4 matches found in sequence:
aaa98926 ; H. tuberculata hemocyanin Hth2 domain a'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98926 check: 5557 from: 1 to: 591
ID AAA98926 standard; DNA; 591 BP.
XX
AC AAA98926;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth2 domain a'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX

853 TATACTATGACAACTTTGAAATCAGTGGTTTAAACTTAAATGAGATCGAGCGGTTAATA
903 908
913 GCMAAACGCAAGTCACATGCTAGAGTCTTTGCTGGGTCTCTGTGT
...
1156 CCATTTTCTGAGGCTGTCGGTTGTTGCTGTGAATGGAACCTGTCATTCGATCGTCTCAT
1206 1211
1216 CTTCCACGCAACGATAATCTATGAACCGCGAA
1227 1232

8 matches found in sequence:
aaa98924 ; H. tuberculata hemocyanin HthI domain g'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98924 check: 1514 from: 1 to: 1209
ID AAA98924 standard; DNA; 1209 BP.
XX AC AAA98924;
XX DT
XX DE
XX H.. tuberculata hemocyanin HthI domain g'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
PA
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
PS Claim 1a; Page 153; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate

CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1209 BP; 313 A; 337 C; 285 G; 274 T; 0 other;
AAA98924 Length: 1209 March 5, 2002 14:18 Type: N Check: 1514 ..
Found using 'seq2-3' (pappu403.key)
1 GATCACCATGACGACCATCATCTAGTCGGGAAGCATAGCAGGATCCGGGGTCCGGAAGACGCTG
10 15
61 AACACCTTGACTAAGGCTGAGACCGCACAACTGAGGGAGCGCTGTGGGTGTCTCATGCGCA
99 104
121 GACCACGGTCCCAATGGCTTTCAAGCTATTGCTG
...
195 CCACAACCTACTCATGTTTGTACTACGGCATGGCTACCTTCCACACTGGCATCGCTCTA
245 250
255 CACCAAGCAGATGGAGGATGCAATGAGGGCGCATGGGTCTCATGTC
...
392 TTGATTATCTCAATGTGACGACAACTCGATCTCCCGAGACATGCTGTTCAACGACGCCCG
442 447
452 AGCATGGATCAGATGCTGTTTCTTACAGACAAAGTCTCTTAGCTCTGGAAACAACTGATT
465 470
512 TCTGCAAAAT
...
685 TACAGAGGACTTCCATACAACCATGCTCAATTTGTGAGATCCAGGCAATGAAACGCGCCCTG
735 740
745 AGGCCTTTCTAGTGACGATATCAACCAACCAACCCAGTCACAAAGGCTAAGCGGAAGCATTA
757 762
805 GATGTGTT
...
1100 ACCTGAGGCATGACTCTGACTTTTACCTTCAGGGTGAAGATTGTGGGACCCAGCCACG
1150 1155
1160 AGCTTCTTTCAGACAGTGTCAAGCACCACCACTATTGAATTGAAC
...

8 matches found in sequence:
aaa98925 ; H. tuberculata hemocyanin HthI domain h'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98925 check: 2420 from: 1 to: 1536
ID AAA98925 standard; DNA; 1536 BP.
XX AC AAA98925;
XX DT 16-FEB-2001 (first entry)
XX

| | | |
|----------|---|-------------|
| PA | (BIOS-) BIOSVN ARZNEIMITTEL GMBH. | |
| XX | | |
| PI | Markl J, Altenhein B, Lieb B, Stiefel T; | |
| XX | | |
| DR | WPI; 2000-587517/55. | |
| XX | | |
| PT | New nucleic acid encoding hemocyanin, useful for gene therapy of tumors | |
| PT | and for recombinant production of fusion proteins for vaccination - | |
| XX | | |
| PS | Claim 1a; Page 152; 163pp; German. | |
| XX | | |
| CC | This invention describes a novel nucleic acid (I) containing a sequence | |
| CC | that encodes hemocyanin (II), a domain of (I) or its fragment with the | |
| CC | immunological properties of at least one domain of (II). The products of | |
| CC | the invention have cytostatic, virucide, antibacterial, antiparasitic, | |
| CC | immunomodulatory and antihypertensive activity. (I), and constructs | |
| CC | additionally containing antigen-encoding sequences, are useful in gene | |
| CC | therapy of tumors. Polypeptides encoded by (I) are useful for treating | |
| CC | parasitic or viral infections and tumors, particularly schistosomiasis | |
| CC | and carcinoma (of bladder, epithelium, ovary, breast, bronchi or | |
| CC | colon-rectum), also hypertension, as vaccines, for treating cocaine | |
| CC | misuse and very generally as carriers for pharmaceuticals, e.g. | |
| CC | cytostatics. They may also be used to generate antibodies (Ab). Probes | |
| CC | based on (I) and Ab are useful for detecting tumor-specific DNA in a cell | |
| CC | (by detecting specific binding to cellular DNA or proteins), particular | |
| CC | where associated with the types of carcinoma listed above. Hemocyanins | |
| CC | can be produced recombinantly, relatively inexpensively and in adequate | |
| CC | amounts, eliminating the need to culture gastropods. When used as a | |
| CC | carrier, (II) significantly increases the half-life of the attached | |
| CC | pharmaceutical, by inhibiting ultrafiltration in the kidneys. | |
| XX | | |
| SQ | Sequence 1251 BP; 339 A; 287 C; 286 G; 339 T; 0 other; | |
| AAA98923 | Length: 1251 March 5, 2002 14:18 Type: N Check: 9187 .. | |
| | Found using 'seq2-3' (pappu403.key) | |
| ... | | |
| 104 | TTCAACATGATAATGGGACTGATGGTTATCAAGCTATTGCTGCCTTCATGGCGTTCCTG | 154 159 |
| | | ----- ----- |
| 164 | CGCAGTGCACGACGCATCTGGACGTGAGATCGCCTGTTGTCATCCACGGCATGGCGACGT | 216 221 |
| | | 193 198 |
| | | ----- ----- |
| 224 | TTCTCACTGGCACCGGTTGTACACTTCGAGTTGGAGCAGCGCTCGCAGACACGGGT | 264 269 |
| | | ----- ----- |
| 284 | CCAGTGTGCTGTTCCCACTACTGGACTGGACCAAGCCAATCACCAGAACTGCCACACATTC | |
| | | ----- ----- |
| 344 | TGACAGACGGAGAAATATATGACGTTTGGGAAAAATGCCGTCTTGCCCAATCCGTTTGCAA | 364 369 |
| | | ----- ----- |
| 404 | GAGGTTATGTGAAAAAT | |
| ... | | |
| 543 | CGAAGTTTCAGTTTGAAGTGATGCATAACACAGATCCATTATCTCGTAGGAGGCGGTCAAC | 593 598 |
| | | ----- ----- |
| 603 | GTAGCCCTTCCTCCTCTCTCGAGTATTCTCTATACGATCCAAATCTTCT | |
| ... | | |

668 CCG

...

2 matches found in sequence:
aaa98921 : H. tuberculata hemocyanin HthI domain d'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98921 check: 108 from: 1 to: 1239

ID AAA98921 standard; DNA; 1239 BP.
XX
AC AAA98921;
XX
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin HthI domain d'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
XX WO200055192-A2.
XX
PD 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination -
PT
XX
XX Claim 1a; Page 151; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies, e.g.
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1239 BP; 351 A; 297 C; 269 G; 322 T; 0 other;

AAA98921 Length: 1239 March 5, 2002 14:18 Type: N Check: 108
Found using 'seq2-3' (pappu403.key)

...

17 ATAACACTGTTGAAGAAGTTACTGGGCCAGTCATATCAGGAAGAATTGTGACGACCTCA
67 72 |-----|
77 ATACCCGAGAAATGGAAGCCTTAGAGCTGCTTTCTTCGCATATTCAGGACGACGGAACAT
124 129 |-----|
137 ATGAATCTATTGCCCCAGTACCATTGGCAACCAACGAGCAAAATGTCA
...

5 matches found in sequence:
aaa98922 : H. tuberculata hemocyanin HthI domain e'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98922 check: 4704 from: 1 to: 1260
ID AAA98922 standard; DNA; 1260 BP.
XX
AC AAA98922;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin HthI domain e'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
XX WO200055192-A2.
XX
PD 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination -
PT
XX
XX Claim 1a; Page 152; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies, e.g.
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
CC

CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1266 BP; 328 A; 350 C; 294 G; 294 T; 0 other;

AAA98918 Length: 1266 March 5, 2002 14:18 Type: N Check: 5057 ..
Found using 'seq2-3' (pappu403.key)

```
1      CTTGTTCACTTCTACTCGTCCCTGTGTGGTGGGGCTGGAGCAGACACACGTCGTCAG
      19 24      |-----| |-----| |-----|
      49 54 57      |-----| |-----| |-----|
61      AAGGACGTGAGTCACCTCAGCGATCAGCAGGTGCAAGCTCTCCACGGCGCCTCCATGAC
      106 111 118      |-----| |-----| |-----|
121     --|      GTCACTGCATCTACAGGCGCTCTGAGTTTCTGAAGACATAAACATCTTACCATCCGCACCA
      123 180      |-----| |-----| |-----|
181     ----|      GCGTGTGTGACTACAAGGACGGAAGATCGCTGCTGTGTCCACGGTATGCCAGTTTC
      185 208 213      |-----| |-----| |-----|
241     CCGTTCTGSCACAGGCATATGCTGTCCAAAGCGGAGCGGCGACTGTTGTCCAAACGGAAG
      262 267      |-----| |-----| |-----|
301     ACTGTGGAATGCCTTACTGGGACTGGACGCAACGCTGACTCACTTACCATCTCTGTG
      333 338      |-----| |-----| |-----|
361     ACTGAACCATCTACATTACAGTAAGGTGGAAGGCTCAAACCACTACTGTTACCGC
      429 434      |-----| |-----| |-----|
421     GCGGAGATAGCGTTTCATCAATAAGAAGACTGCGCGAGCTGTAGATGATCGCTATTTCGAG
      467 472      |-----| |-----| |-----|
481     AAGTGGAGCCCTGGTCACTACACACATCTTATGGAGACTGTCTCGACGCTCTCGAACAG
      526 531      |-----| |-----| |-----|
541     GACGAATCTGTAAATTTGAAATCCAGTTCGAGTTGGCTCA
      579 584      |-----| |-----| |-----|
629     CAAACTTGAATFACACTCTCTACGACCCCATCTTCTCTCCACCACCTCAACAGTTGACC
      679 684      |-----| |-----| |-----|
689     GCCTCTCGCCATCTGGCAGCGCTCTTCAGGAACCTCGGAGGAAGAATCCCAATGCAATGG
      707 712      |-----| |-----| |-----|
749     ACTGTGCACATGAA
      900 905      |-----| |-----| |-----|
850     GATTACAAACAACTTGGATACAGCTACGACAGACTTAAACCTGAATGAATGACGCAGAA
      932 937      |-----| |-----| |-----|
910     CAGTGAAACAGAACTAGACGACGCCACTCCAAAGAACGTGCGTTTGAAGCTTCCGA
      994 999      |-----| |-----| |-----|
```

```
970      CTCAGTGGCTTTGGGGTCTTGCCAACTGTGTGTCTATGTCATGTGTCCCTGATGATGAT
      994 999      |-----| |-----| |-----|
1030     CCACGCGAGTGACTACTGCGAGAAAGCAGCGCACTTCTTCAATCTTGGGGGTCAAAGC
      1060 1065      |-----| |-----| |-----|
1090     GAAATGCCGTGGAGATTCTACAGACC
      1257 1257      |-----| |-----| |-----|
...
8 matches found in sequence:
aaa98919 ; H. tuberculata hemocyanin Hth1 domain b'' encoding DNA.
(from "mycobacterieng.seq")
TOIG of: aaa98919 check: 5519 from: 1 to: 1257

ID AAA98919 standard; DNA; 1257 BP.
XX
AC AAA98919;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain b'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
XX
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
DR WPI: 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
PS Claim 1a: Page 150; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
```

3 matches found in sequence:
aaa98917 ; H. tuberculata hemocyanin KLH2 domain g' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98917 check: 1349 from: 1 to: 1185

ID AAA98917 standard; DNA; 1185 BP.
XX
AC AAA98917;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin KLH2 domain g' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
PS Claim 1a; Page 126; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1185 BP; 319 A; 283 C; 270 G; 312 T; 1 other;
AAA98917 Length: 1185 March 5, 2002 14:18 Type: N Check: 1349
Found using 'seq2-3' (pappu403.key)

1 ATCATATTGCTGGCAGTGCAGTACGGAAGACGTGAGCTCTTCCATCCGACATCTGAGATG
35 40
I-----I

61 AGAACCTGAGGCATGCTCTGCAAGCGTGA

...
124 TTGCTGCTTATACGGAAGTCTCCCATGTGTGCATGTCNTGATGGTAGACACGTTGCAT
174 179
I-----I

184 GTTGTACTCATGGAATGGCATCTTTCCCTCACTGGCACAGACTGTT
...
871 AGAGAAAGAGAGAGAGAGAACATTTGCAGCCTTCTCTTGCACGGATTTGGCGCCAGTG
921 926
I-----I

931 CTGATGTTTCTGTTGATGCTGTGCACACCTGATGGTCAATTTGTGCCTT
...

18 matches found in sequence:
aaa98918 ; H. tuberculata hemocyanin Hth1 domain a'' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98918 check: 5057 from: 1 to: 1266

ID AAA98918 standard; DNA; 1266 BP.
XX
XX AC AAA98918;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain a'' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
XX OS Haliotis tuberculata.
XX OS
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
PS Claim 1a; Page 150; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell

CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.

XX Sequence 1257 BP; 331 A; 305 C; 290 G; 331 T; 0 other;
AAA98908 Length: 1257 March 5, 2002 14:18 Type: N Check: 8091 ..
Found using 'seq2-3' (pappu403.key)

...
481 TCAAGGTGGACAGCTTTTGTATAAAGTGGTTCTTCAACAAGCCATCCTAGCGCTGAGC
531 536
|-----|

541 AGGAAACTACTGTGACTTTGAGATTTCAGTTTGAAATCTTCAACAGCGGCTTCACACGT
588 593
|-----|

601 GGGTCGGAGCGAGTCGTACTCTCTATCGGACATCTTCATTA

...
748 CTCACTGTCTCTCGAGCAATGAGAGACCATTTGAAGCCTTTTCAGCTTCGGCGCTCCTT
798 803
|-----|

808 ATAACGGATCAGCTCACACAGGATTTCTCCGACCCGAGGACAC

...

5 matches found in sequence:
aaa98916 ; H. tuberculata hemocyanin KLH2 domain encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98916 check: 3614 from: 1 to: 1251
ID AAA98916 standard; DNA: 1251 BP.
XX AC AAA98916;
XX DT 16-FEB-2001 (first entry)
XX DE H. tuberculata hemocyanin KLH2 domain encoding DNA.
XX KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX OS Haliotis tuberculata.
XX PN WO20005192-A2.
XX PD 21-SEP-2000.
XX PF 17-MAR-2000; 2000WO-EP02410.
XX PR 17-MAR-1999; 99DE-1011971.
XX PR 20-AUG-1999; 99DE-1039578.
XX PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.

XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX Claim 1a; Page 125; 163pp; German.
PS This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX Sequence 1251 BP; 339 A; 296 C; 267 G; 349 T; 0 other;

AAA98916 Length: 1251 March 5, 2002 14:18 Type: N Check: 3614 ..
Found using 'seq2-3' (pappu403.key)

...
61 CCGAGAGAGATCGCCAGCCTGAATCTGCAATGAGGTCTCTACAAGCTGACGATGGG
111 116
|-----|

121 TGAACGGTTATCAAGCCATTGTCATCTCCAGCGTCTCCCGGCTTC

...
373 GGGATGCAGTTGTCAATAATCCATTGTCTAAAGGCTACATTAATCCGAGGACGCTTACA
423 428
|-----|

433 CGGTTAGGGATCCTCAGGACATTTTGTACCATTGCGAGGAGGAAAC

...
508 TTTTAGCCTTAGACACACAGATTTCGTGATTTTGAGGTTCAATTTGAGTCTGCTCATTA
558 563
|-----|

568 ATGCTATTCACTACTTGGTGGTGGTGGCGACAAGTTTATGTCTCTTC

...
662 CTTTGTGACAAAATATGGCAGCTCTGGCAAGCTCTGCAAAAGAGAGAGCGTCCCTA
712 717
|-----|

722 TCATAAAGCGGATTGTCTCTTAACATGATGACCAACCAATGCGA

...
958 ATGGCATTTGGAAGATCAGCTGATGTACGATTTTGGATTTTCAAGACAGCTGACGCTGCC
1008 1013
|-----|

1018 ACGCATCTGGCATGATCTTTTATCTTAGGAGGTTCTTAAAGAGATGCA

...
1018 ACGCATCTGGCATGATCTTTTATCTTAGGAGGTTCTTAAAGAGATGCA

XX Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
XX pharmaceutical carrier; ds.
OS Halotis tuberculata.
XX WO200055192-A2.
XX 21-SEP-2000.
XX 17-MAR-2000; 2000WO-EP02410.
XX 17-MAR-1999; 99DE-1011971.
XX 20-AUG-1999; 99DE-1039578.
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
XX Claim 1a; Page 121; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
XX Sequence 1242 BP; 339 A; 296 C; 275 G; 332 T; 0 other;
SQ
AAA98907 Length: 1242 March 5, 2002 14:18 Type: N Check: 8246
Found using 'seq2-3' (pappu403.key)
...
73 TGTCAAGGAGAGTAGAGAGCCTAAGGTCTCCCTTCTCGCAACTTCAGACGACGGAG
123 128
133 TCTATGAGATAATTGCCAAGTTCACGGCAAGCCTGGGTGTGTGA
...
330 GAGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCATTCCCGTCAACAACGTTTGAC
380 385
390 CCTAATCCTTCTTTCAGAGGTAAATCAGTTTGTGAGAAATGCTGTTA
...
436 CAACACGTGATCCCCAGCCTGAGCTGTAGCTTAACAGGTACTACTACCAAAACGTCATGT
486 491

496 TGGTTTTGAACAGGACAACACTACTCGGACTTCGAGATACAGTTTGA
...
863 GTTCAATCATTTTCAGTATTCAGGCTTGAAGAAATCATTCGTATTAGACAGCTCAAGA
913 918
923 TCGTGTGTTTCAGGATTCCTCTTCACAAACATTGGGACATCCGCA
...
1017 TGTCAAAACAAAGCCGGAACATTTTGCCTACTCGGAGGAGAAACAGAGATGCGCTTCAT
1067 1072
1077 TTTGACAGACTCTACAGGTTTGACATCATCAGTGAAACACTGAGGAGACC
...

3 matches found in sequence:
aaa98908 ; H. tuberculata hemocyanin Hth1 domain e' encoding DNA.
(from "mycobacterng_seq")
TOIG of: aaa98908 check: 8091 from: 1 to: 1257
ID AAA98908 standard; DNA; 1257 BP.
XX
XX AC AAA98908;
XX
XX DT 16-FEB-2001 (first entry)
XX
XX DE H. tuberculata hemocyanin Hth1 domain e' encoding DNA.
XX
XX KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
XX OS Halotis tuberculata.
XX
XX WO200055192-A2.
XX
XX PD 21-SEP-2000.
XX
XX PF 17-MAR-2000; 2000WO-EP02410.
XX
XX PR 17-MAR-1999; 99DE-1011971.
XX
XX PR 20-AUG-1999; 99DE-1039578.
XX
XX PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
XX Claim 1a; Page 121-122; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine

```
167 TTTTCGAGAAGTTGGACCAGAGAGAAATACCGGACTCTTTGAAGGAATT 161 166
...
227 TTGAACAGGATGAATTCTGCACTTCGAGATCCAGTTTGAGTTGGCTCAACAGCGCTATCC 277 282
287 ACTACCTGGTTGGCGCGCTCACACGTACTCCATGTCATCTCATCTCGA
...
354 CTCTTCTCTCCATCACTCCAAACGCGGACCGCATCTTCGCCATCTGGGAACGTCTTCA 404 409
414 GGTACTCAGAGGAAGGACCCCAACACGCGGCTCGGCACACAAC
...
7 matches found in sequence:
aaa98906 ; H. tuberculata hemocyanin Hth1 domain c' encoding DNA.
(from "mycobacterieng.seq")
TOIG of: aaa98906 check: 5951 from: 1 to: 1246

ID AAA98906 standard; DNA; 1246 BP.
XX
AC AAA98906;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain c' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
DR WPI; 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
PS Claim 1a; Page 120-121; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
```

```
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1246 BP; 315 A; 323 C; 299 G; 309 T; 0 other;

AAA98906 Length: 1246 March 5, 2002 14:18 Type: N Check: 5951
Found using 'seq2-3' (pappu403.key)
...
175 TGTCCATTCCCGAGGCCACAAATAGTTGCGCTGTTCATCCACGCGCATTC 225 230
|-----|
235 CCTCATTTGGCACAGACTGTTGTCACCCAGGTGGAAGATGCTCTGTATCAGCGGAGGATCG 291
|-----|
295 CCTATAGGGGTCCCTACTTGGGACTGGACTCAGCCTATGGCGCATCTCCAGGACTTGCA 296
|-----|
355 GACAAGCGCCACCTATAGAGATCCCATCAGCGGGGACAGACAGACACAACCCCTTCCAGGAT 358 363
|-----|
415 GTTGAAGTTGCCCTTTGAAATGGACGTACAGAACGTCACCCAGCATAGATTGTTTGAA 446 451
|-----|
475 CAACCTTTATTTTGGCAAAACATACGCGT
...
652 ATCTTCTACCTTCGTCACTCCAACACTGACCGGCTCTGGGCAATTTGGCAAGGTTGCAG 702 707
|-----|
712 ATACGAAGAACAAGGCCTTACAAAGGCTCATTTGCTTGGTCTGAGGAACCCAGCCTCTC 758 763
|-----|
772 AAACCTTTGCGCTTCAGTTCCCTCCACTGAACAACAACGAAAAAACCTACGAAAACTCGGTG
|-----|
832 CCCACCAACGTTTACGACTACGAAGGAGTCCCTTGGCTATATCTTATGATGACCTCAACTTC 838 843
892 GG
...
5 matches found in sequence:
aaa98907 ; H. tuberculata hemocyanin Hth1 domain d' encoding DNA.
(from "mycobacterieng.seq")
TOIG of: aaa98907 check: 8246 from: 1 to: 1242

ID AAA98907 standard; DNA; 1242 BP.
XX
AC AAA98907;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain d' encoding DNA.
```


1 GGCTTTGTTCACTTCTTCTGCTCGCCTTGTGGTGGGCTGGAGCAGACACGCTCGTCA
21 26 51 56 54 59
|-----|
|-----|
|-----|
61 GAAAGGACGTGAGTCACCTCAGGATGACGAGGTGCAAGCTCTCCACGGCGCCCTCCATG
108 113 120
|-----|
|-----|
121 ACGTCACTGCATCTACAGGGCCTCTGAGTTTCGAAGACATAACATCTTACCATGCCGCAC
125
|-----|
|-----|
181 CAGCGTGTGTGACTACAAGGACGGAAGATCGCTGCTGTGTCCACGGTATGCCAGTT
182 187 210 215
|-----|
|-----|
241 TCCCTTCTGCCACAGGCGATATGTGTCTCAAGCCGAGCGGGCACACTGTTGTCCAAACGGA
264 269
|-----|
|-----|
301 AGACTGTCGGAATGCCTTACTGGACTGGACGCAACCGCTGACTCACTTACCATCTCTTG
335 340
|-----|
|-----|
361 TGA CTGAACCATCTACATTCACAGTAAGGTGGAAGGCTCAAAACCAACTACTGGTACC
|-----|
|-----|
421 GCGGAGATAGCTTCTATCAATAAAGACTGCGCGAGCTGTAGATGACGCTATTCG
431 436 469 474
|-----|
|-----|
481 AGAAGGTGGAGCGCTGCTCACTACACATCTTATGGAGACTGCTCTCGACGCTCTCGAAC
528 533
|-----|
|-----|
541 AGGACGAATCTGTAAATTTGAATCCAGTTTCGAGTTGGCTCA
|-----|
|-----|
631 CAAACTTGGATACACCTCTACAGCCCATCTTCTTCTCCACCACCTCCAAAGCTTGACC
681 686
|-----|
|-----|
691 GCCTCTTCCGCATCTGGCAGGCTCTTCAGGAACCTCGAGGAAAGAAATCCCAATGCAATGG
709 714
|-----|
|-----|
751 ACTGTGCACATGAA
|-----|
|-----|
852 GATTACAAACAACTTGGATACAGCTACGACAGCTTAACCTGAATGGAATGACGCCAGAA
902 907
|-----|
|-----|
912 CAGCTGAAACAGAACTAGACGACGCGCACTCCAAAGAACGTCGCTTTGCAAGCTTCGGA
934 939
|-----|
|-----|
972 CTCAGTGGCTTTGGGGTTCTGCCAACGTTCTGTCTATGATGATGTCCCTGATGATGAT
996 1001
|-----|
|-----|
1032 CCACGCACTGATGACTACTGCGAAGAACGAGGACACTTCTTCATTTCTTGGGGGTCAAAGC
1062 1067
|-----|
|-----|
1092 GAAATGCGGTGGAGATCTACAGACC
|-----|
|-----|

4 matches found in sequence:
aaa98905 ; H. tuberculata hemocyanin HthI domain b' encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98905 check: 6585 from: 1 to: 569
ID AAA98905 standard; DNA; 569 BP.
XX
AC AAA98905;
XX
16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin HthI domain b' encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
pharmaceutical carrier; ds.
XX
OS Hallotis tuberculata.
XX
PN WO200055192-A2.
XX
21-SEP-2000.
XX
17-MAR-2000; 2000WO-EP02410.
XX
17-MAR-1999; 99DE-1011971.
XX
20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
WPI; 2000-587517/55.
XX
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
and for recombinant production of fusion proteins for vaccination -
XX
PS Claim 1a; Page 120; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
that encodes hemocyanin (II), a domain of (I) or its fragment with the
immunological properties of at least one domain of (II). The products of
the invention have cytostatic, virucide, antibacterial, antiparasitic,
immunomodulatory and antihypertensive activity. (I), and constructs
additionally containing antigen-encoding sequences, are useful in gene
therapy of tumors. Polypeptides encoded by (I) are useful for treating
parasitic or viral infections and tumors, particularly schistosomiasis
and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
colon-rectum), also hypertension, as vaccines, for treating cocaine
misuse and very generally as carriers for pharmaceuticals, e.g.
cytostatics. They may also be used to generate antibodies (Ab). Probes
based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
by detecting specific binding to cellular DNA or proteins), particularly
where associated with the types of carcinoma listed above. Hemocyanins
can be produced recombinantly, relatively inexpensively and in adequate
amounts, eliminating the need to culture gastropods. When used as a
carrier, (II) significantly increases the half-life of the attached
pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 569 BP; 142 A; 174 C; 126 G; 127 T; 0 other;
AAA98905 Length: 569 March 5, 2002 14:18 Type: N Check: 6585 ..
Found using 'seq2-3' (pappu403.key)
...
107 GGTATCGTGGAAACATCAAGTTTGAGATATAAGAGACTGCAAGAGCTTTGACGATCCGC
157 162
|-----|
|-----|

PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination
PS Claim 1a; Page 83; 163pp; German.
XX
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
XX pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1546 BP; 438 A; 401 C; 329 G; 377 T; 1 other;

AAA98894 Length: 1546 March 5, 2002 14:18 Type: N Check: 270 ..
Found using 'seq2-3' (pappu403.key)

...

90 TGTCCCTGGTGAAGCAACAACCTTGAGAAATGCCCTTTACAAGCTACAGAACGACACACA
140 145

150 GTCTAACGGGATACGAAGCAATCTCTGGTTACCATGGATACCCCAATCTGTGTCGGGAAG

210 AAGCGGATGACAAATACCCCTGCTGGCGTCCCGGATGGGCATCTTCCCTTACTGGCA
212 217

...

305 CACAATGGTGCACCTGTGGTGTCTCTACTGGGACTGGAACAGGACCTGTCTGCTACTG
355 360

365 CCGCGGTTCTTCTCCGACTCCAGCAACAACAATCCCTACTTTCAAGTACCACATCGCGGT
367 372

425 GTTGGTCACGACACCGTCAGAGAGCCCAACTAGTCTTATATATAACCA

...

522 TTGAAGAAACAATTACTGGGACTTTGAGGTTTCAGTAGATGAGATCCTCCACAACGCGTCC
572 577

582 ACTCCTGGCTGGAGGATCCAGAGAATATTCCATGTCTACCCCTGGA

...

719 AAGCCCTACAACTTCGCTAAATGTGCTTATCATATGATGGAAGACGACCTGGCCCTTC
769 774

779 AGCTATCCATCTATCAACCAGGACGAGTTTACCCGTCGCCAACTCCA

...

AAA98904 Length: 1269 March 5, 2002 14:18 Type: N Check: 2449 ..
Found using 'seq2-3' (pappu403.key)

1254 TGTCTTGTATTATCCAGTANGAAGAAACCTTCACATCCCTCCCAAAAGTTGTGTCACAGA
1304 1309

1314 AAGCACCCGCATCGAGTTCCACCCAGTCGATGATTCAAGTTACGAG

...

18 matches found in sequence:

aaa98904 : H. tuberculata hemocyanin Hth1 domain a' encoding DNA.
(from "mycobactereng.seq")
TOIG of: aaa98904 check: 2449 from: 1 to: 1269

ID AAA98904 standard; DNA; 1269 BP.

XX AAA98904;

AC AAA98904;

DT 16-FEB-2001 (first entry)

XX H. tuberculata hemocyanin Hth1 domain a' encoding DNA.

DE H. tuberculata hemocyanin Hth1 domain a' encoding DNA.

XX Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;

KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;

KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;

KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;

KW pharmaceutical carrier; ds.

XX Haliotis tuberculata.

OS WO200055192-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-EP02410.

PF 17-MAR-1999; 99DE-1011971.

XX 20-AUG-1999; 99DE-1039578.

PR (BIOS-) BIOSYN ARZNEIMITTEL GMBH.

XX Markl J, Altenhein B, Lieb B, Stiefel T;

XX WPI; 2000-587517/55.

XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors

PT and for recombinant production of fusion proteins for vaccination

XX Claim 1a; Page 120; 163pp; German.

XX This invention describes a novel nucleic acid (I) containing a sequence

CC that encodes hemocyanin (II), a domain of (I) or its fragment with the

CC immunological properties of at least one domain of (II). The products of

CC the invention have cytostatic, virucide, antibacterial, antiparasitic,

CC immunomodulatory and antihypertensive activity. (I), and constructs

CC additionally containing antigen-encoding sequences, are useful in gene

CC therapy of tumors. Polypeptides encoded by (I) are useful for treating

CC parasitic or viral infections and tumors, particularly schistosomiasis

CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or

CC colon-rectum), also hypertension, as vaccines, for treating cocaine

```

-----
6 matches found in sequence:
aaa98893 ; H. tuberculata hemocyanin Hth2 domain g encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98893 check: 5461 from: 1 to: 1207

ID AAA98893 standard; DNA; 1207 BP.
XX
AC
XX
XX
DT 16-FEB-2001 (first entry)
XX
DE
XX
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS
XX
XX
PN WO200055192-A2.
XX
XX
PD 21-SEP-2000.
XX
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX
DR WPI; 2000-587517/55.
XX
XX
New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
and for recombinant production of fusion proteins for vaccination -
XX
XX
Claim la; Page 82-83; 163pp; German.
XX
XX
This invention describes a novel nucleic acid (I) containing a sequence
that encodes hemocyanin (II), a domain of (I) or its fragment with the
immunological properties of at least one domain of (II). The products of
the invention have cytostatic, virucide, antibacterial, antiparasitic,
immunomodulatory and antihypertensive activity. (I), and constructs
additionally containing antigen-encoding sequences, are useful in gene
therapy of tumors. Polypeptides encoded by (I) are useful for treating
parasitic or viral infections and tumors, particularly schistosomiasis
and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
colon-rectum), also hypertension, as vaccines, for treating cocaine
misuse and very generally as carriers for pharmaceuticals, e.g.
cytostatics. They may also be used to generate antibodies (Ab). Probes
based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
(by detecting specific binding to cellular DNA or proteins), particularly
where associated with the types of carcinoma listed above. Hemocyanins
can be produced recombinantly, relatively inexpensively and in adequate
amounts, eliminating the need to culture gastropods. When used as a
carrier, (II) significantly increases the half-life of the attached
pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1207 BP; 302 A; 335 C; 283 G; 287 T; 0 other;

AAA98893 Length: 1207 March 5, 2002 14:18 Type: N Check: 5461
Found using 'seq2-3' (pappu403.key)

...
CTGTGACAGTGCGCACTCTGCCAACATTCGTGGCTCTGGGGTGAAGAGCACTCACGA
57 62
|-----|

```

```

67 CCTCACTGTGTCTGTAGAGACCGAGAACCTAAGACAGGCTCTTTCAAGG
...
391 TTGATCATCTCGGTGTAACACAGTCACGTTCCCCCAGAGACATGCTGTTTAAAGACCCAG
|-----|
441 446
451 AGCAAGGATCAGAGTCGTTCTTCTATAGACAAGTCCTCTGGCTTTGGAGCAGACTGACT
|-----|
464 469
511 ACTGCCAGTTTCAAGTCCAGTTTGAGCTGACCCACACGCCATTTCATCTCTGGACAGGTG
|-----|
546 551
571 GACGTAGCCCTTACGGAATGTCGACCCCTCGA
...
703 ACGAAGCACACTGTGAATCCAGGTTCTGAAACAGCCCTTGAGGCCATTTCACAGATGACA
|-----|
753 758
763 TCAACACATCCAAATCACCAGAGACTAATGCCAGGCCTATCGATTG
...
910 AGAGAACATTGCTGCTTCCTCTCTGCGAATCGGTGTCAGTCTGCTCTTTG
|-----|
960 965
970 ACATCTCGCGCCCAATGGTGACTGTGTCTTTGCAAGAACCTTTGC
...
-----
8 matches found in sequence:
aaa98894 ; H. tuberculata hemocyanin Hth2 domain h encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98894 check: 270 from: 1 to: 1546

ID AAA98894 standard; DNA; 1546 BP.
XX
AC
XX
XX
DT 16-FEB-2001 (first entry)
XX
DE
XX
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS
XX
XX
PN WO200055192-A2.
XX
XX
PD 21-SEP-2000.
XX
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX
DR WPI; 2000-587517/55.
XX
XX

```

```
PS Claim la; Page 81-82; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
XX Sequence 1255.BP; 331 A; 305 C; 289 G; 330 T; 0 other;
SQ
AAA98891 Length: 1255 March 5, 2002 14:18 Type: N Check: 900
Found using 'seq2-3' (pappu403.key)
...
479 TCAAAGTGGACACAGTTTGTGATAAACTGGTTCTTCAACAAGCCATCCTAGCGCTGAGC
|-----|
529 534

539 AGGAAACTACTGTGACTTTGAGATTGAGTTTGAAATCTTCACACGGCGTTACACAGT
|-----|
586 591

599 GGGTCGGAGCAGTCGTACCTACTCTATCGGACATCTTCATTA
|-----|
796 801

746 CTCACGTGCTCTCGAGCAATGAGAGAACCAATTGAAGCCTTTTCAGCTTCGGCGCTCCTT
|-----|
796 801

806 ATAACTGGAATCAGCTCACACAGGATTTCTCCCGACCCGAGGACAC
|-----|
...

-----
4 matches found in sequence:
aaa98892 ; H. tuberculata hemocyanin Hth2 domain f encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98892 check: 5919 from: 1 to: 1248

ID AAA98892 standard; DNA; 1248 BP.
XX
AC AAA98892:
XX
XX 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth2 domain f encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
XX immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
```

```
PD 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EF02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
PI WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
XX
XX Claim la; Page 82; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
XX Sequence 1248 BP; 326 A; 312 C; 291 G; 319 T; 0 other;
SQ
AAA98892 Length: 1248 March 5, 2002 14:18 Type: N Check: 5919
Found using 'seq2-3' (pappu403.key)
...
1 GTGACATAAATACCAGGAGCATGTCACCGAACCGTGTTCGCCGTGAGCTGAGCGATCTGT
|-----|
51 56

61 CTGCGAGGGACCTGTCTAGTCTCAAGTCTGCTCTGCGAGACCTACA
|-----|

...

238 GACTGTACACCCCTGCAGTTGGAGATGGCTCTGAGGAGACATGGATCATCTGTGCGCATCC
|-----|
288 293

298 CCTACTGGGACTGGACAAAGCCCTATCTCCGAACCTCCCTCGCTCTT
|-----|

...

539 TGAGGTACAGTTTGGAGTCTCCCATACGTGATCCACTACTCTTGGTGGAGCTCAGAC
|-----|
589 594

599 CTACGCATTGTCTTCTCTGCAATTATGCTCTCTACGACCCCATTTCTTC
|-----|

...

1152 CCATTTTATATCAAGGTTGAGATCCATGCTGTTAAACAAGACCATGATACCGTCGTGTTG
|-----|
1202 1207

1212 ATCCCAAGCCCAACTATCATCTATTCTCTGGGGAAG
```

5 matches found in sequence:
aaa98890 ; H. tuberculata hemocyanin Hth2 domain d encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98890 check: 5460 from: 1 to: 1244

ID AAA98890 standard; DNA; 1244 BP.
XX
AC AAA98890;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth2 domain d encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
XX WO200055192-A2.
XX
PN 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
XX
PR 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination
XX
XX Claim 1a; Page 81; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
XX that encodes hemocyanin (II), a domain of (I) or its fragment with the
XX immunological properties of at least one domain of (II). The products of
XX the invention have cytostatic, virucide, antibacterial, antiparasitic,
XX immunomodulatory and antihypertensive activity. (I), and constructs
XX additionally containing antigen-encoding sequences, are useful in gene
XX therapy of tumors. Polypeptides encoded by (I) are useful for treating
XX parasitic or viral infections and tumors, particularly schistosomiasis
XX and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
XX colon-rectum), also hypertension, as vaccines, for treating cocaine
XX misuse and very generally as carriers for pharmaceuticals, e.g.
XX cytostatics. They may also be used to generate antibodies (Ab). Probes
XX based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
XX (by detecting specific binding to cellular DNA or proteins), particularly
XX where associated with the types of carcinoma listed above. Hemocyanins
XX can be produced recombinantly, relatively inexpensively and in adequate
XX amounts, eliminating the need to culture gastropods. When used as a
XX carrier, (II) significantly increases the half-life of the attached
XX pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1244 BP; 339 A; 296 C; 276 G; 333 T; 0 other;
XX
AAA98890 Length: 1244 March 5, 2002 14:18 Type: N Check: 5460 ..
Found using 'seq2-3' (pappu403.key)

...

73 TGTCRAAGGGAGNAGTAGAGAGCCTAAGGTCTGCCCTTCTCGCAACTTCAGAACGCGGAG
123 128

133 TCTATGAGAATATTGCCAAGTTCACGGCAAGCCTGGTGTGTGA

...
330 GAGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCATTCGCCGTCAACAACGTTTCAC
380 385
390 CCTAATCCTTTCTTCACAGAGTAAATCAGTTTTGAGAAATGCTGTGTA
436 CAACACGTGATCCCGCAGCCTGAGCTGTACGTTAACAGGTACTACTACCAAAACGTCATGT
486 491
496 TGGTTTTTGAACAGAGCAACTACTTCCGACTTCGAGATACAGTTTGA
...
863 GTTCAATCATTTACAGTATTCACAGGCTTGAAGAAATCATTCGTATTAGACAAACGTCAGA
913 918
923 TCGTGTGTTTGCAGGATTCCTCTTCACAAACATTGGGACATCCGCA
...
1017 TGTGAAAAACAAGCCGGAACATTTCCGCTACTCGGAGGAGAAACAGAGATGCCGTTTCAT
1067 1072
1077 TTTCACAGACTCTACAGGTTTGACATCAGTCAACACTGAGGAGACC
...

3 matches found in sequence:
aaa98891 ; H. tuberculata hemocyanin Hth2 domain e encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98891 check: 900 from: 1 to: 1255

ID AAA98891 standard; DNA; 1255 BP.
XX
AC AAA98891;
XX
XX 16-FEB-2001 (first entry)
DT
XX
XX H. tuberculata hemocyanin Hth2 domain e encoding DNA.
XX
XX Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
XX Haliotis tuberculata.
OS
XX WO200055192-A2.
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
XX
XX 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination
XX
XX

```
XX SQ Sequence 1003 BP; 263 A; 249 C; 233 G; 258 T; 0 other;
AAA98888 Length: 1003 March 5, 2002 14:18 Type: N Check: 8886
Found using 'seq2-3' (pappu403.key)

1 CACAGACTGTTGCTCACCCAGGTGGAAGATGCTCTGTGATCAGGCGAGGATCGCTATAGGG
48 53 |-----|
61 GTCCCCTACTGGGACTGGACTCAGCCTATGGCGCATCTCCAGGACTTGCAGACAACGCC
115 120 |-----|
121 ACCTATAGAGATCCCATCAGCGGGACAGCAGACACAACCCCTTCCACGATGTTGAAGTT
203 208 |-----|
181 GCCTTTGAAATGGAGCTACAGAACGTCAACCCAGATAGTAGATTGTTTGAAACACCTTTA
241 TTTGGCAACATACGCGT
...
409 ATCTTCTACCTTCGTCACCTCCAACACTGACCGGCTCTGGGCAATTTGGCAAGGTTGCAG
459 464 |-----|
469 ATACGAAGAACAAGCGCTTACAAAGGCTCATTTGTTGCTGTGAGAACGCCAGCCTCTC
515 520 |-----|
529 AAACCTTTGCGCTTCAGTTCCCTCCACTGAACACACGAAACAACTACGAAACTCGGTG
589 CCCACCAACGTTTACGACTACGAAGGAGTCCTTTGGCTATATCTATGATGACCTCAACTTC
595 600 |-----|
649 GG
...
5 matches found in sequence:
aaa98889; H. tuberculata hemocyanin Hth2 domain c encoding DNA.
(from "mycobacterieng.seq")
TOIG of: aaa98889 check: 7964 from: 1 to: 1251

ID AAA98889 standard; DNA; 1251 BP.
XX AC AAA98889;
XX DT 16-FEB-2001 (first entry)
XX DE H. tuberculata hemocyanin Hth2 domain c encoding DNA.
XX KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
XX KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
XX KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
XX KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
XX KW pharmaceutical carrier; ds.
XX OS Halliotis tuberculata.
XX PN WO200055192-A2.
XX PD 21-SEP-2000.
XX PF 17-MAR-2000; 2000WO-EP02410.
XX PR 17-MAR-1999; 99DE-1011971.
XX PR 20-AUG-1999; 99DE-1039578.
XX
```

```
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX WPI; 2000-587517/35.
DR New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination
PT Claim 1a; Page 80-81; 163pp; German..
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX SQ Sequence 1251 BP; 332 A; 310 C; 281 G; 328 T; 0 other;
AAA98889 Length: 1251 March 5, 2002 14:18 Type: N Check: 7964
Found using 'seq2-3' (pappu403.key)
...
397 ACCCTGGTACAGCGCCATATTGATACAGTTGGTGTGACACACAAGAGCGTCGCT
447 452 |-----|
457 AAGAACTGTATGAAGCTCCTGGATTGGCCATTATATCTGGGGTCGCTAAGCAAGTGCTTC
498 503 |-----|
517 TGGCTTTGGAGCAGGATGACTTCTGTGATTTTGAAGT
...
567 GCTCACAATTTCATTCACGCTCTTGTGCGGGGAAGCGCCATATGTTATGGCGTCACTC
617 622 |-----|
627 CGTTACACTACTTATGATCCAAATTTTCTACCTCCATCAATTCTAACA
...
1043 TGGTGAATTTCTACATCCTTTGGTGATGAAGCTGAATGCCATGGGGCTATGATCGTCTTTA
1093 1098 |-----|
1103 CAAATATGAGATCACTGAGCAGCTCAATGCCCTGGATCTACACATCGGAGATAGATCTT
1175 1180 |-----|
1163 CATCAGATACGAAGCGTTTGGATCTTTCATGTTACAGTCTTTGGAAGCAACATCTTCCCAA
1223 ACCTTCTG
...
-----
```



```

XX Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haletotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-EP02410.
XX
XX 17-MAR-1999; 99DE-1011971.
XX
XX 20-AUG-1999; 99DE-1039578.
XX
XX (BIOS-) BIOSYN AR2NEIMITTEL GMBH.
XX
XX Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX and for recombinant production of fusion proteins for vaccination -
XX
XX Claim 1a; Page 79; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
XX that encodes hemocyanin (II), a domain of (I) or its fragment with the
XX immunological properties of at least one domain of (II). The products of
XX the invention have cytostatic, virucide, antibacterial, antiparasitic,
XX immunomodulatory and antihypertensive activity. (I), and constructs
XX additionally containing antigen-encoding sequences, are useful in gene
XX therapy of tumors. Polypeptides encoded by (I) are useful for treating
XX parasitic or viral infections and tumors, particularly schistosomiasis
XX and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
XX colon-rectum), also hypertension, as vaccines, for treating cocaine
XX misuse and very generally as carriers for pharmaceuticals, e.g.
XX cytostatics. They may also be used to generate antibodies (Ab). Probes
XX based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
XX (by detecting specific binding to cellular DNA or proteins), particularly
XX where associated with the types of carcinoma listed above. Hemocyanins
XX can be produced recombinantly, relatively inexpensively and in adequate
XX amounts, eliminating the need to culture gastropods. When used as a
XX carrier, (II) significantly increases the half-life of the attached
XX pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
XX Sequence 1209 BP; 313 A; 337 C; 285 G; 274 T; 0 other;
XX
XX
XX A9A9886 Length: 1209 March 5, 2002 14:18 Type: N Check: 1748 ..
XX Found using 'seq2-3' (pappu403.key)
XX
1 ATCCACATGACGACCATCAGTCGGGAAGCATAGCAGGATCCGGGGTCCGCAAGCAGCTGA
9 14
1
61 ACACCTTGACTAAGGCTGAGACCGACACACCTGAGGGAGGCGCTGTGGGGTGTGCATGCAG
98 103
121 ACCACGGTCCCATGGCTTTCAGCTATTGCTG
...
194 CCACAACACTACTACTGTTGTACTACGCGCATGCGTACCTTCCACACATGGCATCCGCTCTA
|-----|
244 249
254 CACCAACGAGATGGAGGATGCAATGAGGCGCATGGGTCTCATGTC
..

```


917 TTCTGGCATTGATGAACCTTATCCAGCAGACAGGAGAAACACAGA 907 912

...

1179 GTAACATACCGTCTTACAAATGGACAGACTACTCTCGGAGACCTCATTCAGACCCCTCC 1229 1234

1239 ATTATATTGTACTGGACGCC 1255 1260

10 matches found in sequence:
 aaa98885 ; H. tuberculata hemocyanin Hth1 domain f encoding DNA.
 (from "mycobacterng.seq")
 TOIG of: aaa98885 check: 8076 from: 1 to: 1251

ID AAA98885 standard; DNA; 1251 BP.
 XX
 AC AAA98885;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE H. tuberculata hemocyanin Hth1 domain f encoding DNA.
 XX
 KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
 KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
 KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
 KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
 KW pharmaceutical carrier; ds.
 XX
 OS Haliotis tuberculata.
 XX
 PN WO200055192-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-EP02410.
 XX
 PR 17-MAR-1999; 99DE-1011971.
 PR 20-AUG-1999; 99DE-1039578.
 XX
 PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
 XX
 PI Markl J, Altenhein B, Lieb B, Stiefel T;
 XX
 DR WPI; 2000-587517/55.
 XX
 PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
 PT and for recombinant production of fusion proteins for vaccination
 XX
 PS Claim 1a; Page 78-79; 163pp; German.
 XX
 CC This invention describes a novel nucleic acid (I) containing a sequence
 CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
 CC immunological properties of at least one domain of (II). The products of
 CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
 CC immunomodulatory and antihypertensive activity. (I), and constructs
 CC additionally containing antigen-encoding sequences, are useful in gene
 CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
 CC parasitic or viral infections and tumors, particularly schistosomiasis
 CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
 CC colon-rectum), also hypertension, as vaccines, for treating cocaine
 CC misuse and very generally as carriers for pharmaceuticals, e.g.
 CC cytostatics. They may also be used to generate antibodies (Ab). Probes
 CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
 CC (by detecting specific binding to cellular DNA or proteins), particularly
 CC where associated with the types of carcinoma listed above. Hemocyanins
 CC can be produced recombinantly, relatively inexpensively and in adequate
 CC amounts, eliminating the need to culture gastropods. When used as a

CC carrier, (II) significantly increases the half-life of the attached
 CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
 XX
 SQ Sequence 1251 BP; 339 A; 286 C; 287 G; 339 T; 0 other;
 AAA98885 Length: 1251 March 5, 2002 14:18 Type: N Check: 8076
 Found using 'seq2-3' (pappu403.key)
 ...

103 TTCACATGATAATGGACTGATGTTATCAAGCTATTGTCCTTCATGGCTTCTCTG 153 158

163 CGCAGTCCACGAGCCCATCTGGAGCTGGAGATCGCCTGTGTCATCCACGGCATCGGACGT 192 197
 215 220
 218

223 TTCCTCACTGGCACCGCTTGTACACTCTGCAGTTGGAGCAAGCGCTGCGCAGACACGGGT 263 268

283 CCAGTGTGCTGTTCATCTACTGGACTGGACCAAGCAATCACCGAACTGCCACACATTC

343 TGACAGACGGAGATATATTACAGCTTTGGCAAAATGCCGCTTGGCCCAATCCGTTTGCAA 363 368

403 GAGGTTATGTGAAAT

542 CGAAGTTTCAGTTTGAAGTGATGCATAACACGATCCATTATCTCTAGAGGGCGTCAAC 592 597

602 GTACGCCTTCTCCTCTCTCGAGTATTCCTCATACGATCCATCTTC

852 TATAACTATGACAACTTGAAATCAGTGGTTTAAACTTAAATGAGATCGAGCGCTTAATA 902 907

912 GCAAAACGCAAGTCACATGCTAGAGCTTTGCTGGGTTCTCTGTGT

1155 CCATTTTCTCGAGGCTGTCGGTTGTTGCTGTGAATGGAACGTGTCATTCCATCGCTTCAT 1205 1210

1215 CTTCAACGAGCAACGATAATCTATGAACACGGCGAAG 1226 1231

8 matches found in sequence:
 aaa98886 ; H. tuberculata hemocyanin Hth1 domain g encoding DNA.
 (from "mycobacterng.seq")
 TOIG of: aaa98886 check: 1748 from: 1 to: 1209

ID AAA98886 standard; DNA; 1209 BP.
 XX
 AC AAA98886;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE H. tuberculata hemocyanin Hth1 domain g encoding DNA.

```
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
DR WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination -
XX
PS Claim 1a; Page 77-78; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1239 BP; 351 A; 297 C; 269 G; 322 T; 0 other;
AAA98883 Length: 1239 March 5, 2002 14:18 Type: N Check: 8874
Found using 'seq2-3' (pappu403.key)
...
16 ATAACCTACGTTGAAGAAGTTACTGGGCCAGTCATATCAGGAAGATTTGACGACCTCA
|-----|
66 71
76 ATACCGGAGAAATGGAAAGCCTTAGAGCTGCTTTCCTGCATATTCAGGACGACGGAAACAT
|-----|
123 128
136 ATGAATCTATTGCCCCAGTACCATGGCAACACGAGGCAAAATGTCA
...
-----
6 matches found in sequence:
aaa98884 ; H. tuberculata hemocyanin Hth1 domain e encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98884 check: 2013 from: 1 to: 1260
ID AAA98884 standard; DNA; 1260 BP.
XX
AC AAA98884;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain e encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
```

```
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
DR WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination -
XX
PS Claim 1a; Page 78; 163pp; German.
XX
XX This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1260 BP; 353 A; 288 C; 281 G; 338 T; 0 other;
AAA98884 Length: 1260 March 5, 2002 14:18 Type: N Check: 2013
Found using 'seq2-3' (pappu403.key)
...
68 CCCAAGGGAGAGGGTTTCTTAGTCAAGCTTTTGCAAGAAGATGAAGAATGATCGTCCCG
|-----|
118 123
128 TGATGGTACCAGCATTGCCTCTTCCATGCCCTGCCACCACCTC
...
388 TTAATATTTCAAATCCATTCCTCGGGGCTGACATAGATTTGAAGGACGCGGCTTCATA
|-----|
438 443
448 CAGAGAGGCACATAAATACTGAGCGCCTGTTTTCACAGTGGGATCATGACGGATACCAACA
|-----|
469 474
508 ACTGGTCTTCGAAACT
...
857 GTTGACTATAAAGAAGTTGGTGGATACAGATATGATAGTCTGTAATTGGAGGGCGATCAAT
|-----|
```

```
638 GCATTACACAGCCTTTGACCCCAATTTTCTACCTTCATCAATTCCAATGTGATCCTCTATG 688 693
698 GGCAATCTGGCAAGCTCTTCAAATCAGGAGACACAAGCCATATCAA
...
843 ACTGACATTTATGACTACGAGGAAGTGTGCTGACTACAGCTACAGTATCAACGTTTGGT 893 898
903 GGGATGAACCTTGAAGAAATAGAGAAGCTATACATCTCAGACAA
...
1041 CTCAAAGCTGGAGATATTGCCATTTCTTGTTGGTGCCAAAGAAATGCTTGGGGCTTTGAC 1091 1096
1101 CGCTTGTAAGGTGCGAAATAACTGACTCATTTGAGACACTTTCTC
...
1150 ATGTCGATGGAGATTATGAAGTCACITTTTAAATTCATGATGCCACGGAACGCTCTTG 1200 1205
1210 ATACGGACCTGATTCCACACGACGACGCTGTTGTTTCTGAGCCAGCTCA
...
-----
3 matches found in sequence:
aaa98882 ; H. tuberculata hemocyanin Hth1 domain c encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98882 check: 1428 from: 1 to: 1242

ID AAA98882 standard; DNA; 1242 BP.
XX
AC AAA98882;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain c encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-EP02410.
XX
PR 17-MAR-1999; 99DE-1011971.
PR 20-AUG-1999; 99DE-1039578.
XX
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX
PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX
XX WPI; 2000-587517/55.
XX
XX New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
PT and for recombinant production of fusion proteins for vaccination
PS
PS Claim la; Page 77; 163pp; German.
XX
```

```
CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX
SQ Sequence 1242 BP; 354 A; 282 C; 271 G; 335 T; 0 other;

AAA98882 Length: 1242 March 5, 2002 14:18 Type: N Check: 1428
Found using 'seq2-3' (pappu403.key)
...
88 CCCTGGAGCTTCTTGAATAATGATCATCTACTGCAGGTGGATTCAATCAGCTTGCGCCCTTCC 138 143
148 ATGGAGAGCCTAAATGTGGCCCTAATCTGAAGCGGAGCACAAAGGT
...
547 TGCAGTATGAGATTTCCTCCATAATTTTATCCATGCATCTGTAGGAGGAACCCAGCTATG 597 602
607 GCATGGCATCGCTGAGATATACAGCATACGATCCAATCTTTTCTTGCATCATTCAAACA 614 619
667 CCG
...
-----
2 matches found in sequence:
aaa98883 ; H. tuberculata hemocyanin Hth1 domain d encoding DNA.
(from "mycobacterng.seq")
TOIG of: aaa98883 check: 8874 from: 1 to: 1239

ID AAA98883 standard; DNA; 1239 BP.
XX
AC AAA98883;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain d encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
XX
OS Haliotis tuberculata.
XX
PN WO200055192-A2.
XX
PD 21-SEP-2000.
XX
```

```
241 TCCCTTCTGCGCAGGCGCATATGTCGTCCAAAGCCGAGCGGCACTGTTGTCCAAACGGA
      |-----|
      264 269
301 AGACTGTCGGAATGCCTTACTGGGACTGGAGCGCAACGCTGACTCACTTACCATCTCTTG
      |-----|
      335 340
361 TGACTGAACCCATCTACATTGACAGTAAAGGTGGAAGGCTCAAAACCACTACTGGTACC
      |-----|
      431 436
421 GCGGCGAGATAGCGTTTCATCAATAAGAAGACTGCGGAGCTGTAGATGATCGGCTATTTCG
      |-----|
      469 474
481 AGAAGGTGGAGCCTGCTCACTACACACATCTTATGGAGACTGTCTCTCGACGCTCTCGAAC
      |-----|
      528 533
541 AGGACGAATTCTGTAAATTTGAATCCAGTTCGAGTTGGCTCA
...
631 CAAACTTGGAAATACACCTCTCTAGACCCCATCTTCTTCTCCACCACTCAACGTTGACC
      |-----|
      681 686
691 GCCTCTTCGCCATCTGCGACGGTCTTCAGGAAGTCCGAGGAAGAAATCCCAATCAATGG
      |-----|
      709 714
751 ACTGTGCACATGAA
...
852 GATTACAAACAACCTTGGATACAGTACGACAGCTTAAACCTGAATGGAATGACGCGCAGAA
      |-----|
      902 907
912 CAGCTGAAACAGAACTAGACGACGCGCACTCCAAAGAACGTGCGTTTGCAAGCTTCCGA
      |-----|
      934 939
972 CTCAGTGGCTTTGGGGTTCTGCGCAACGTTGTGTCTATGATGCTGTGCCCTGATGATGAT
      |-----|
      996 1001
1032 CCACGCGAGTGATGACTACTCGGAGAAAGCGCGGACTTCTTCATCTTTGGGGTCAAAGC
      |-----|
      1062 1067
1092 GAAATGCCGTGGAGATTCTACAGACC
...
-----
8 matches found in sequence:
aaa98881: H. tuberculata hemocyanin Hth1 domain b encoding DNA.
(from "mycobacterieng.seq")
TOIG of: aaa98881 check: 3801 from: 1 to: 1257

ID AAA98881 standard; DNA; 1257 BP.
XX
AC
AC AAA98881;
XX
DT 16-FEB-2001 (first entry)
XX
DE H. tuberculata hemocyanin Hth1 domain b encoding DNA.
XX
KW Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW pharmaceutical carrier; ds.
```

```
XX OS Haliotis tuberculata.
XX PN WO200055192-A2.
XX PD 21-SEP-2000.
XX PF 17-MAR-2000; 2000WO-EP02410.
XX PR 17-MAR-1999; 99DE-1011971.
XX PR 20-AUG-1999; 99DE-1039578.
XX PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX PF; 2000-587517/55.
XX PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX PT and for recombinant production of fusion proteins for vaccination
XX PS Claim 1a; Page 76-77; 163pp; German.
XX CC This invention describes a novel nucleic acid (I) containing a sequence
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
CC immunological properties of at least one domain of (II). The products of
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
CC immunomodulatory and antihypertensive activity. (I), and constructs
CC additionally containing antigen-encoding sequences, are useful in gene
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
CC parasitic or viral infections and tumors, particularly schistosomiasis
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
CC colon-rectum), also hypertension, as vaccines, for treating cocaine
CC misuse and very generally as carriers for pharmaceuticals, e.g.
CC cytostatics. They may also be used to generate antibodies (Ab). Probes
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
CC (by detecting specific binding to cellular DNA or proteins), particularly
CC where associated with the types of carcinoma listed above. Hemocyanins
CC can be produced recombinantly, relatively inexpensively and in adequate
CC amounts, eliminating the need to culture gastropods. When used as a
CC carrier, (II) significantly increases the half-life of the attached
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX SQ Sequence 1257 BP; 336 A; 305 C; 280 G; 336 T; 0 other;
AA98881 Length: 1257 March 5, 2002 14:18 Type: N Check: 3801
Found using 'seq2-3' (pappu403.key)
1 GTCACCTTGACCCACCTGTGCATCATCGCCACGATGACGATCTTATTGTTTCGAAAAATA
      |-----|
      25 30 35 41
61 TAGATCATTTGACTCGTGTGAAGAGGAAATACGA
...
246 TGGCACC GGCTGTGTTACCCAGGTGGAAGATGCTCTTGTACGGCGTGATGCCCTATC
      |-----|
      296 301
306 GGTGTTCTCTTATTGGGACTGGACAAAACCTATGACTCACCTTCACG
...
436 AGGACACCATCACACGAGCAGGATGATAGATCGAAACTGTTTGGCCCCAGTCGCTTTTG
      |-----|
      486 491
496 GGGAGCATTCCTCCATCTGTTTGATGGAATCCTGTACCGCATTTGACCA
...

```

```

559 AGCAGCGGACGGTGTGCGGACCAACCAAGATCACCGGAACTCGTCTGCTGACGACATCG |-----|
619 CGACGCTGGCGGCTCAGCTGTGACCTCCGAGAGACGTCAAAACAGGTGCGCCACACCGTCT |-----|
679 GGATCGCTCGGATGGATCCTCGCACCTGGTCCAGATCCAGATCGCTCCACCAAGACA |-----|
739 CGTCGGTGCAGTTGACGATGTCGACTGCGGTGGGTAGCAGGTCAACCGCCACCAACCGGTCT |-----|
799 AGCATCGTCTAATTTGTTGAAACCGCGCACCGCGGCATTCGGAAGACTCAC |-----|
859 CAATGACGACGATGATTACTCTTCGCGGACGGTTGCGGTTGCGGTCGCGGCGTCCCA |-----|
919 CTGCGCGCGGACGACCGTCAACCTGGCTCCCGCACCAAGTTCGCGGATGCTCATG |-----|
979 GCGCAA
...
1063 AGTCGTGTGGCTACTCGACTGCAAGTGTCTCACCAGTTTCACCGGCTGGCGGCGGTGCG |-----|
1123 CGCCACGATAGGCATACCGAGGAGAGTTGGACCCACCTTGGCGCGCGCCATGAAGG |-----|
1183 ACGCCTGACCAAGCTCGGCGGCGCTACATCGACACCTGGGCTGCAACTATCCGCA |-----|
1243 CGCCCGCTTAGCGAGCGGATCGGTATCCGCGGTGCATCCCGGTAGCAACGCTGT |-----|
1303 CGACTCGCGCGGACAGATCGGCTCGCCGCGACCGGTGAACCTCGCG |-----|
...
18 matches found in sequence:
aaa98880 ; H. tuberculata hemocyanin Hth1 domain a encoding DNA.
TOIG of: aaa98880 check: 1871 from: 1 to: 1269
ID AAA98880 standard; DNA; 1269 BP.
XX
AC AAA98880;
XX
DT 16-FEB-2001 (first entry)
XX

```

```

DE XX H. tuberculata hemocyanin Hth1 domain a encoding DNA.
KW XX Hemocyanin; cytostatic; virucide; antibacterial; antiparasitic;
KW XX immunomodulatory; antihypertensive; gene therapy; tumor; treatment;
KW XX infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;
KW XX breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;
KW XX pharmaceutical carrier; ds.
XX OS Haliotis tuberculata.
XX XX WO200055192-A2.
XX PN 21-SEP-2000.
XX PD
XX PF 17-MAR-2000; 2000WO-EP02410.
XX PR 17-MAR-1999; 99DE-1011971.
XX PR 20-AUG-1999; 99DE-1039578.
XX XX
XX PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.
XX PI Markl J, Altenhein B, Lieb B, Stiefel T;
XX PT WPI; 2000-587517/55.
XX DR
XX XX
XX PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors
XX PT and for recombinant production of fusion proteins for vaccination
XX PS Claim 1a; Page 76; 163pp; German.
XX CC
XX CC This invention describes a novel nucleic acid (I) containing a sequence
XX CC that encodes hemocyanin (II), a domain of (I) or its fragment with the
XX CC immunological properties of at least one domain of (II). The products of
XX CC the invention have cytostatic, virucide, antibacterial, antiparasitic,
XX CC immunomodulatory and antihypertensive activity. (I), and constructs
XX CC additionally containing antigen-encoding sequences, are useful in gene
XX CC therapy of tumors. Polypeptides encoded by (I) are useful for treating
XX CC parasitic or viral infections and tumors, particularly schistosomiasis
XX CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or
XX CC colon-rectum), also hypertension, as vaccines, for treating cocaine
XX CC misuse and very generally as carriers for pharmaceuticals, e.g.
XX CC cytostatics. They may also be used to generate antibodies (Ab). Probes
XX CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell
XX CC (by detecting specific binding to cellular DNA or proteins), particularly
XX CC where associated with the types of carcinoma listed above. Hemocyanins
XX CC can be produced recombinantly, relatively inexpensively and in adequate
XX CC amounts, eliminating the need to culture gastropods. When used as a
XX CC carrier, (II) significantly increases the half-life of the attached
XX CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.
XX SQ Sequence 1269 BP; 328 A; 351 C; 297 G; 293 T; 0 other;

AAA98880 Length: 1269 March 5, 2002 14:18 Type: N Check: 1871
Found using 'seq-3' (pappu403.key)

1 GGCTGTTTCAGTTTCTACTCGTCGCCCTTTGTGGCGGGGCTGGAGCAGACAAGCTGCTCA |-----|
21 26 51 56 59

61 GAAGACGTGAGTCACTCACGGATGACGAGGTGCAAGCTCTCCACGGCGCGCTCCATG |-----|
108 113 120

121 ACCTCATGTCATACAGGGCCTCTGAGTTTCGAAGACATACATCTTACCATGCCGCAC |-----|
125

181 CAGCGTCGTGTGACTACAAGGACGGAAGATGCCCTGTGTGTCCACGGTATGCCCGAGTT |-----|
182 187 210 215

```

```

85      CCGTCGATCGCGTCTCTCGCGCCGTTAGCATCGCGCACTGTGTTGCGCGGTGCTC
      92  97      |-----| |-----| |-----|
145      GTGGGCTCGAAGCCAAAGCGGGGACCACTTCGGGACGC
      ...
233      GGCACATGGTCTGACGGTCAAGGCAAGATCCCGGACTGTCTCTGAAGAGCTGAGC
      283 288      |-----|
293      GCGATCTCACCACCAACCCACCGCGCGACGGGAAACGTCAAGCTCACGCTGGGTGG
      293 298      |-----| |-----| |-----|
353      TCTGATATCGATCGCGACTTCGTGTGTTTCGACGGGATCTGTACGCCACCCCTGAGCGCC
      406 411      |-----|
413      AACAGTGGAGGATTTTCGTCCTCCGCCGCGGACATCTAGACCCGCCCGGAGTGAAT
      422 427      |-----|
473      CCGGA
      ...
551      AACACCATCGCATCAGCGGAAGGTATCGGCACAGCGGTGAACCATAGAGCGCGCG
      601 606      |-----|
611      TTCAACGCGACGACCGGTGCGCGGACCGTCTGGATTTCAGGAGACCGCGGATCAATCAA
      634 639      |-----|
671      CTGGCACAGGCCAGTTGGACCGCGCTCGGGCAATTCCGTCCA
      ...
738      AGAAGTCCAGGTCACGAAGCCCGCGTGAGCTGATCAGCATCGGACGAGCGTGGAGT
      788 793      |-----|
798      CCGATTAGCGCG
      -----
29 matches found in sequence:
aaa53603 ; M. tuberculosis antigen 23 coding sequence.
(from "mycobacter.ng.seq")
TOIG of: aaa53603 check: 1651 from: 1 to: 1380

ID    AAA53603 standard; DNA; 1380 BP.
XX
AC    AAA53603;
XX
DT    31-OCT-2000 (first entry)
XX
DE    M. tuberculosis antigen 23 coding sequence.
XX
KW    Interleukin 12; IL-12; gamma-interferon; antigen; stimulate;
KW    immunostimulant; immunomodulatory; secretory protein; ss.
XX
OS    Mycobacterium tuberculosis.
XX
FH    Key Location/Qualifiers
FT    CDS 66..800
FT    /tag=a
FT    /product= Antigen_23
FT    /note= "A. M. tuberculosis secreted protein"
XX
PN    WO200039301-A2.
```

```

XX      06-JUL-2000.
PD
XX      23-DEC-1999; 99WO-US30975.
PF
XX      24-DEC-1998; 98US-0220416.
PR
XX      (CORI-) CORIXA CORP.
PA
XX      Skeiky Y;
PI
XX      WPI; 2000-452399/39.
DR      P-PSDB; AAY96928.
DR
XX      Potentiating immune responses to antigens using specified Mycobacterium
XX      tuberculosis proteins
XX
PS      Example 6.2; Fig 6; 30pp; English.
XX
CC      The Mycobacterium tuberculosis proteins encoded by AAA53599-603
CC      stimulate interleukin (IL)-12 and gamma-interferon production. A number
CC      of cloned M. tuberculosis antigens were expressed as recombinant proteins
CC      in Escherichia coli. After purification, these proteins were tested for
CC      their ability to stimulate the production of IL-12 by a mouse macrophage
CC      cell line, RAW264. The concentrations of IL-12 in RAW264 culture
CC      supernatants were measured by an ELISA using an antibody to detect the
CC      p40 component of the IL-12. Among the 18 proteins tested, only 2,
CC      designated DPV and DPAS stimulated the production of IL-12. The
CC      stimulatory activities of these proteins were not affected by the
CC      inclusion of polymixin B in the cultures, which abrogated the IL-12
CC      stimulatory activities of lipopolysaccharide, a polyclonal B cell
CC      mitogen. Both DPV and DPAS are secretory proteins of M. tuberculosis.
CC      A method of using the M. tuberculosis proteins to induce and potentiate
CC      immune reactions to antigens is claimed.
XX
SQ      Sequence 1380 BP; 276 A; 479 C; 396 G; 229 T; 0 other;
      AAA53603 Length: 1380 March 5, 2002 14:18 Type: N Check: 1651 ..
      Found using 'seq2-3' (pappu403.key)
      ...
67      TGAAGCATCCACCTTGTTCGGTGTTCGGCGCCGCCACCGCATACTCGCGTGTGCTGG
      117 122      |-----| |-----| |-----|
127      CGATCGGCGGTGCTCAACCGAAGGGGACGCGGCAAGCGTCTGTACACCGCGGCACCG
      130 153 158 164 169      |-----|
187      CATCCAAACGGCGATGCGGCCATGCTACTCAAGCAGGCCACCGATGCGATGCGCAAGGTCA
      195 200      |-----|
247      CCGAATGCACGTCAGACTTGGGTGACAGCGGCGCGCATGCGCAAAACCTGCGGGTGACCAAGC
      276 281      |-----|
307      TCGAAGCGGATATCTCAACACACACCGCAGACGCGTTGCCACCGGTAGCGGACATTGCTCG
      312 317      |-----|
367      T
      ...
412      ACCTGTACTCCGACCTAGCGCGGCGCACCTACACCGATTTCGGCAACGCGGCGCTCGA
      462 467      |-----|
472      TCTACAACGCTGTCGGTGTCTTCGACCCCAACAAGGGCGCTGGCCAA
      ...
```

CC designated DPV and DPAS stimulated the production of IL-12. The
CC stimulatory activities of these proteins were not affected by the
CC inclusion of polymixin B in the cultures, which abrogated the IL-12
CC stimulatory activities of lipopolysaccharide, a polyclonal B cell
CC mitogen. Both DPV and DPAS are secretory proteins of M. tuberculosis.
CC A method of using the M. tuberculosis proteins to induce and potentiate
CC immune reactions to antigens is claimed.
XX
SQ Sequence 930 BP; 166 A; 312 C; 298 G; 154 T; 0 other;

AAA53601 Length: 930 March 5, 2002 14:18 Type: N Check: 6330 ..
Found using 'seq2-3' (pappu403.key)

...
16 CACCATCACTTTCCCGCGCGGCTTCGCGGTGGAGTACCTGAGGTCGCGTCGCCGTGCG
66 71
76 ATGGCGCGTGACATCAAGTCCAAATTCAAAGTGGTGCCAACTGCGCGCGCGCTGTAC
136 CTGCTCGACGCGCTGCGCGCGAGGAGCTTTCAGCGGCTGGGACATCAACACCGCGCGG
160 165
196 -|
TTCGAGTGTACGACCATCGCGGCTTCGCGTGTATCGCGGTGGTGGCC
197
...
332 GCGAGCTGCGGGGTGGCTGCAGGCCAACAGGACGTCAGGCCACCGGAAGCGCGCTCG
382 387
388
-|
TCGGTCTTCGATGGCTGCTTCTCGCGGTGAGCTGGCGATCTATCAACCCCGCAGAGT
417 422 429 434
393 423 428
452 TCCTCTACGGGAGCGATGTCGGGCTGTGTGACCCCTCCAGCGGATGGTCCACCC
465 470
512 TGATCGCGCTGGGGATGGGTGAGCTGGCGGCTACAGGCTCCGACATGTGGGGCCCGA
522 527 532 537
572 AGGAGACCGCGGTGGCAGCGCAGCACCGCGCTGTGTGAACGTCGGGAGCTGATGCCA
595 600 610 615
632 ACAACACCGCGTCTGGGTGTACTGCGGCAAGCGGACGCGTGGATCT
698 CGGCCAGGTTCTTCGAGGCTTCGTCGGGACCAACATCAAGTTCAGAGCGCTACA
748 753 757
----|
762
758 AGCGCGTGGCGGCCAACAGGCGTGTTCGACTTCCCGACAGCGGTAGCACAGCTGGG
818 AGTACTGGGGCGGCAGCTCAACGCTATGAAGCCCGACCTGCAACGGGCACTGGGTGCCA
838 843
878 CGCCCAACACCGGGCGCGCGCGCTAGCTCCGGAACAGACGAAATTC
-----|

25 ACCGGCAACTCACAGACCCCTCTACGATGCAGGGTATGCGGACCCCGACGCCACTG
75 80
|-----|

13 matches found in sequence:
aaa53602 ; M. tuberculosis antigen 22B coding sequence.
(from "mycobacterng.seq")
TOIG of: aaa53602 check: 9023 from: 1 to: 810

ID AAA53602 standard; DNA; 810 BP.

XX AC AAA53602;

XX AC 31-OCT-2000 (first entry)

XX DE M. tuberculosis antigen 22B coding sequence.

XX KW Interleukin 12; IL-12; gamma-interferon; antigen; stimulate;
XX KW immunostimulant; immunomodulatory; secretory protein; ss.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers
XX CDS 53..772

XX FT /*tag= a

XX FT /product= Antigen_22B

XX FT /note= "A M. tuberculosis secreted protein"

XX PN WO200039301-A2.

XX PD 06-JUL-2000.

XX PF 23-DEC-1999; 99WO-US30975.

XX PR 24-DEC-1998; 98US-0220416.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y;

XX DR WPI; 2000-452399/39.

XX DR P-PSDB; AAY96926.

XX PT Potentiating immune responses to antigens using specified Mycobacterium
XX PT tuberculosis proteins

XX PS Example 6.2; Fig 5; 30pp; English.

XX CC The Mycobacterium tuberculosis proteins encoded by AAA53599-603
XX CC stimulate interleukin (IL)-12 and gamma-interferon production. A number
XX CC of cloned M. tuberculosis antigens were expressed as recombinant proteins
XX CC in Escherichia coli. After purification, these proteins were tested for
XX CC their ability to stimulate the production of IL-12 by a mouse macrophage
XX CC cell line, RAW264. The concentrations of IL-12 in RAW264 culture
XX CC supernatants were measured by an ELISA using an antibody to detect the
XX CC p40 component of the IL-12. Among the 18 proteins tested, only 2,
XX CC designated DPV and DPAS stimulated the production of IL-12. The
XX CC stimulatory activities of these proteins were not affected by the
XX CC inclusion of polymixin B in the cultures, which abrogated the IL-12
XX CC stimulatory activities of lipopolysaccharide, a polyclonal B cell
XX CC mitogen. Both DPV and DPAS are secretory proteins of M. tuberculosis.
XX CC A method of using the M. tuberculosis proteins to induce and potentiate
XX CC immune reactions to antigens is claimed.

SQ Sequence 810 BP; 165 A; 280 C; 249 G; 116 T; 0 other;

AAA53602 Length: 810 March 5, 2002 14:18 Type: N Check: 9023 ..
Found using 'seq2-3' (pappu403.key)

...

XX The Mycobacterium tuberculosis proteins encoded by AAA53599-603
CC stimulate interleukin (IL)-12 and gamma-interferon production. A number
CC of cloned M. tuberculosis antigens were expressed as recombinant proteins
CC in Escherichia coli. After purification, these proteins were tested for
CC their ability to stimulate the production of IL-12 by a mouse macrophage
CC cell line, RAW264. The concentrations of IL-12 in RAW264 culture
CC supernatants were measured by an ELISA using an antibody to detect the
CC p40 component of the IL-12. Among the 18 proteins tested, only 2,
CC designated DPV and DPAS stimulated the production of IL-12. The
CC stimulatory activities of these proteins were not affected by the
CC inclusion of polymixin B in the cultures, which abrogated the IL-12
CC stimulatory activities of lipopolysaccharide, a polyclonal B cell
CC mitogen. Both DPV and DPAS are secretory proteins of M. tuberculosis.
CC A method of using the M. tuberculosis proteins to induce and potentiate
CC immune reactions to antigens is claimed.

XX Sequence 1237 BP; 224 A; 404 C; 393 G; 216 T; 0 other;

AAA53600 Length: 1237 March 5, 2002 14:18 Type: N Check: 5217 ..
Found using 'seq2-3' (pappu403.key)

...

86 ATCGAGCGGCAACGAACGAAGAACACACACCATGAAGATGGTGAATCGATCGCGCAG
136 141

146 GTCTGACCGCGCGGTGCAATCGCGCGCTGCGCGCGGTGTGACTTCGATCATGGCTG
169 174

206 GCGGCGCGGTGTATACAGATGCAGCGGTGTCTTCGCGCGCGCACTGCGCTGGAGCC
235 240

266 CGGCATCCGCCCTGAGCTCCCGACCGCGCCAGTTGACCAGCGCTGCTCAACAGCCTCG
280 285

326 CCGATCCCAAGCTGCTGTTTGGCAACAGGGCAGTCTGGTCGAGGGCGGCATCGGGGCA
339 344

386 CCGAGCGCGCATCCCGACCAACAGCTGAAGAGCCCGCCGAGCACGGGATCTGCCGC
397 402

446 TGTCTGTTACGGTGTGACGAACATCCAGCGCGCGCGCGGTTCGGCCACCGCGCGGTTT
447 452

506 CCGTCTCGGGTCCGAAGCTCTCGTCCCGGTCACGACAGACGTACAGTTCTGTAATCAAG
528 533

566 GCGGCTGGATGTGTACCGCATCGCGCATGGAGTTGCTGCAGCGCGCGGGAAGTATGAT
591 596

626 TGGCGGCGCGGCTTCAGCCCG

...

729 ACGCACCGCGCGGTGCAAGCCGCTCTCGAGATAGTGTGCTGCCACCGACCGACCC
779 784

789 CCGCGCTCCGCTGTTCTCGCTGCTGCATGAGTTGTCGACAGTGGTGTGTCACACCGAGG

849 GCTTGTGACGTATCGCGCGATGCTCATCCGAGCACCCAGGACCGATCGCTGGAGGCTC
863 868

909 GATGCAACGCGATCGCGCGCTGTGCTGATGCTGCCGACGTGGTGCAGCGCGCGCGG
920 925

969 AGGTATCTGTTTCGACGAGACTCCGCTCTTCGTGGGATTACGAATCGCTGCGCGCGGTG
974 979

1029 CCGCGCATTTTGGGTGAGTGGAGGCGTTCCTGCGCGCGGCTAAGACGCTTGGATCG
1054 1059

1089 TGCCGTCGAGGTCTACCGACAGCGCGCGAGTTTCATCGCGGCTCAGGAGGACGAGG
1110 1115

1149 CCGAACGGGCTCCTCGT

...

19 matches found in sequence:

aaa53601 ; M. tuberculosis antigen 85A coding sequence.
(from "mycobacterng.seq")

TOIG of: aaa53601 check: 6330 from: 1 to: 930

ID AAA53601 standard; DNA; 930 BP.

XX AC AAA53601;

XX DT 31-OCT-2000 (first entry)

XX DE M. tuberculosis antigen 85A coding sequence.

XX KW Interleukin 12; IL-12; gamma-interferon; antigen; stimulate;
immunostimulant; immunomodulatory; secretory protein; ss.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers
FT CDS 4..912

FT FT /*tag= a
FT /product= Antigen_85A

XX PN WO200039301-A2.

XX PD 06-JUL-2000.

XX PF 23-DEC-1999; 99WO-US30975.

XX PR 24-DEC-1998; 98US-0220416.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y;

XX DR WPI; 2000-452399/39.

XX DR P-PSDB; AAY96926.

XX PT Potentiating immune responses to antigens using specified Mycobacterium
tuberculosis proteins

XX PS Example 6.2; Fig 4A; 30pp; English.

XX The Mycobacterium tuberculosis proteins encoded by AAA53599-603
CC stimulate interleukin (IL)-12 and gamma-interferon production. A number
CC of cloned M. tuberculosis antigens were expressed as recombinant proteins
CC in Escherichia coli. After purification, these proteins were tested for
CC their ability to stimulate the production of IL-12 by a mouse macrophage
CC cell line, RAW264. The concentrations of IL-12 in RAW264 culture
CC supernatants were measured by an ELISA using an antibody to detect the
CC p40 component of the IL-12. Among the 18 proteins tested, only 2,


```
181      126 131
      A
...
196      CCTAGCGGAGGACATCACCAGGTCTCAGACGGTTCGGGGCGAACCAGTGGCTGCTCGGG
      246 251
      -|
256      CCGCGGCATTGGCTTCGACTTCACCGAACTGCTGGTCAACGGACTCTCACCACCCCTCA
      257
316      ACCTCAAAATTTGAAGGCCGATGGGGGTTCGCTGATCCGCGAGAGCCCGCGAGCGT
      345 350
      |-----|
376      TGACGACTCCACTCCGCGCGCCGCCCGCTGAGGTGTACCTACTGCAGCGCACCAAG
      376 392 397
      |-----|
436      GCCGCGAGGGCAAGCGGCTCAGCAAGACCACCGGATGGGTCCACCGGCGCTCGTTGAAG
      482 487
      |-----|
496      CCANGGCGTTTCATCACTGTCTGTGGGTGAACCTACGAACAGATCAACGACGACGGCCTAC
      501 506
      |-----|
556      ACATAGCTTCGGCGCCGGAAGCGAGCGGCGGCCACAGCTGTGGC
      540 545
      |-----|
667      ACGGCATTAAACCGCACATCATCGTGGCGCGGGTTCGCCGCGAGTTGGACGCCAAGC
      717 722
      |-----|
727      GCGCCATCAAGCAAGCACCGAACTTCGGGCCAGGCTATAAACTCATTTCTCGGGCTGTA
      790 795 801 806
      |-----|
787      AACACGCTTTCATCGCTGCTGACCTGCAGATTTGGTAATGGCGCGCAGTGTGCGCCGA
      838 843
      |-----|
847      TAACCACTTTGAGAGAAGCAAGCCACGCGCCTGTACCTCGACCGCAGACGGCAAC
      874 879
      |-----|
907      GCTGGGAGATGTCGTGCGCTGCTTAACGACGCGTGAATAGTTTGGGAAGGTGTCCCAT
      918 923 932 937
      |-----|
967      AAATGAGGCTGTCGTGACCGCATTCAGCGCCGTTAGCGCGCTGGCAATGTCGTGA
      977 982
      |-----|
1027      CCCTCGGGCGCGGCTCCGACAGATCCCGTGGACGCGGTATTAACACCAACCTGCA
      1041 1046
      |-----|
```

```
1087      ATTACGGGCAGGTAGTAGCTCGCTCAACGCGACGGATCCGGGGGCTGCCGCACAGTTCA
      1146
      ----|
1147      ACGCTCACCGGTGGCGAGTCTTATTGCGCAATTTCTCTCGCCGACCGCCACC
      1151
      ...
1282      GCTCTGCACAACTATTAAAGCCCATGGGGCCCCCATCCCGGACCCGGCATCTGTCGCG
      1332 1337
      |-----|
1342      GGGCTAGGCCAGATTGCCCGCTCCTCAACGGGCGCATCCCGGACCCGGCATCTGTCGCGA
      1394 1399
      |-----|
      1397
1402      CGGGCTAGGCCAGATTGCCCGCTCCTCAACGGGCGCATCCCGGACCCGGCATCTGTCGCGA
      1402
      |
      ...
      -----
19 matches found in sequence:
aaa53600 ; M. tuberculosis antigen DPAS coding sequence.
(from "mycobacterng.seq")
TOIG of: aaa53600 check: 5217 from: 1 to: 1237

ID      AAA53600 standard; DNA; 1237 BP.
XX
AC      AAA53600;
XX
DT      31-OCT-2000 (first entry)
XX
DE      M. tuberculosis antigen DPAS coding sequence.
XX
KW      Interleukin 12; IL-12; gamma-interferon; antigen; stimulate;
KW      immunostimulant; immunomodulatory; secretory protein; ss.
XX      Mycobacterium tuberculosis.
XX
FH      Key
CDS      Location/Qualifiers
          262..624
          /*tag= a
          /product= DPAS_secreted_protein
          /partial
XX      WO200039301-A2.
XX
PD      06-JUL-2000.
XX
PF      23-DEC-1999; 99WO-US10975.
XX
PR      24-DEC-1998; 98US-0220416.
XX
PA      (CORI-) CORIXA CORP.
XX
PI      Skeiky V;
XX
DR      WPI; 2000-452399/39.
DR      P-PSDB; AAY96925.
XX
PT      Potentiating immune responses to antigens using specified Mycobacterium
PT      tuberculosis proteins
XX
PS      Example 6.2; Fig 2; 30pp; English.
```

ID AAA51918 standard; DNA; 843 BP.
XX AAA51918;
AC
DT
XX 31-OCT-2000 (first entry)
DE T helper cell differentially expressed gene.
XX
XX T helper cell; differential expression; 102 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymometric; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; ss.
XX
OS Mus sp.
XX
XX US6084083-A.
PN
XX 04-JUL-2000.
PD
XX 28-MAR-1997; 97US-0829525.
PF
XX 01-MAR-1996; 96US-0609583.
PR
XX 03-MAR-1995; 95US-0398633.
PR
XX 07-JUN-1995; 95US-0487748.
PR
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX
XX Levinson DA;
PI
XX WPI; 2000-464385/40.
DR
XX
XX New isolated human 200 gene products or polypeptides, useful for
PT treating and diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders
XX
XX Disclosure: Column 121-122; 107pp; English.
XX
XX Genes which are differentially expressed within and among T helper (TH)
CC cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
CC can be used diagnostically or as targets for therapeutic intervention.
CC The polypeptides are useful for treating and diagnosing of immune
CC disorders, especially T lymphocyte-related disorders. These disorders
CC include chronic inflammatory diseases and disorders (e.g. Crohn's
CC disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
CC Grave's disease), or atopic conditions (e.g. asthma and allergy,
CC including allergic rhinitis or food allergies). Also included are
CC certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
CC HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.
XX
XX Sequence 843 BP; 253 A; 185 C; 196 G; 209 T; 0 other;
SQ
AAA51918 Length: 843 March 5, 2002 14:18 Type: N Check: 4217 ..
Found using 'seq2-3' (pappu403.key)
...
209 CTGATGAAGAAATGTGACATATCAGAAATCCAGCAGATACCAGCTAAAGGCGGAFCTCA
259 264
269 ACAAGGAGATGTCCTCTGATCATATAAAGAATGTGACTCTGGATGA
...
514 AATGGACAAAAATTTCCACATGGGTGATGAATTAAGGACTGTGGAGAACGATCAGA
564 569
574 ACTGCTATCCACATTGGAGTGGGAGTCTCTGCTGGGTTGACCCCTGG
...

31 matches found in sequence:
aaa53599 ; M. tuberculosis antigen DPV coding sequence.
(from "mycobacterng.seq")
TOIG of: aaa53599 check: 7958 from: 1 to: 1509
ID AAA53599 standard; DNA; 1509 BP.
XX
AC AAA53599;
XX
DT 31-OCT-2000 (first entry)
XX
DE M. tuberculosis antigen DPV coding sequence.
XX
KW Interleukin 12; IL-12; gamma-interferon; antigen; stimulate;
KW immunostimulant; immunomodulatory; secretory protein; ss.
XX
OS Mycobacterium tuberculosis.
FH
XX Key Location/Qualifiers
FT 1053..1301
FT /*tag= a
FT /product= DPV_secreted_protein
FT /partial
XX
PN WO200039301-A2.
XX
XX 06-JUL-2000.
PD
XX 23-DEC-1999; 99WO-US30975.
PF
XX 24-DEC-1998; 98US-0220416.
PR
XX (CORI-) CORIXA CORP.
PA
XX
PI Skeiky Y;
XX
XX WPI; 2000-452399/39.
DR
XX P-PSDB; AAY96924.
DR
XX Potentiating immune responses to antigens using specified Mycobacterium
PT tuberculosis proteins
PT
XX Example 6.2; Fig 1; 30pp; English.
XX
XX The Mycobacterium tuberculosis proteins encoded by AAA53599-603
CC stimulate interleukin (IL)-12 and gamma-interferon production. A number
CC of cloned M. tuberculosis antigens were expressed as recombinant proteins
CC in Escherichia coli. After purification, these proteins were tested for
CC their ability to stimulate the production of IL-12 by a mouse macrophage
CC cell line, RAW264. The concentrations of IL-12 in RAW264 culture
CC supernatants were measured by an ELISA using an antibody to detect the
CC p40 component of the IL-12. Among the 18 proteins tested, only 2,
CC designated DPV and DPAS stimulated the production of IL-12. The
CC stimulatory activities of these proteins were not affected by the
CC inclusion of polymixin B in the cultures, which abrogated the IL-12
CC stimulatory activities of lipopolysaccharide, a polyclonal B cell
CC mitogen. Both DPV and DPAS are secretory proteins of M. tuberculosis.
CC A method of using the M. tuberculosis proteins to induce and potentiate
CC immune reactions to antigens is claimed.
XX
SQ Sequence 1509 BP; 280 A; 501 C; 470 G; 257 T; 1 other;
AAA53599 Length: 1509 March 5, 2002 14:18 Type: N Check: 7958 ..
Found using 'seq2-3' (pappu403.key)
1 TATCGAATTCGGCAGGAAATTCAGCGAAACCATCCGGTATTTCGACGATTCCTGGCCA
|-----|
47 52
61 AACACGGGTGTCAGGTGCGACTGGGCACCTCGGGTGGCCGCCAGGAGTTGACGGGTACG
121 ACGAGTCTGCTTGGCCACCGGGCGTGGCACCAGCGCATTCGGGCCATCCCGGCGATCGACC
|-----|

SQ Sequence 32 BP; 9 A; 7 C; 7 G; 9 T; 0 other;

AAA51908 Length: 32 March 5, 2002 14:18 Type: N Check: 7788
Found using 'seq2-3' (pappu403.key)

1 AAAAAAAAAATTCGAGCGCTAACACAGAGGTGTC
14 19

1 match found in sequence:

aaa51910 ; Forward primer for murine gene 103 fragment amplification.
(from "mycobacterng.seq")
TOIG of: aaa51910 check: 5916 from: 1 to: 31

ID AAA51910 standard; DNA; 31 BP.

XX AAA51910;

AC AAA51910;

DT 31-OCT-2000 (first entry)

DE Forward primer for murine gene 103 fragment amplification.

XX T helper cell; differential expression; 200 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; primer; ss.

OS Mus sp.

XX US6084083-A.

PN 04-JUL-2000.

PD 28-MAR-1997; 9705-0829525.

PF 01-MAR-1996; 9605-0609583.

PR 03-MAR-1995; 9505-0398633.

PR 07-JUN-1995; 9505-0487748.

XX (MILL-) MILLENNIUM PHARM INC.

PA Levinson DA;

XX WPI; 2000-464385/40.

XX New isolated human 200 gene products or polypeptides, useful for
XX treating and diagnosing immune disorders, especially T helper
XX lymphocyte-related disorders

PS Example; Column 89; 107pp; English.

XX AAA51910-11 were used to amplify the cDNA encoding extracellular domain
XX of the murine 103 gene. The amplified fragment was used to construct an
XX IgG1 fusion protein.

XX Genes which are differentially expressed within and among T helper (TH)
XX cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
XX can be used diagnostically or as targets for therapeutic intervention.
XX The polypeptides are useful for treating and diagnosing of immune
XX disorders, especially T lymphocyte-related disorders. These disorders
XX include chronic inflammatory diseases and disorders (e.g. Crohn's
XX disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
XX Grave's disease), or atopic conditions (e.g. asthma and allergy,
XX including allergic rhinitis or food allergies). Also included are
XX certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
XX HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.

SQ Sequence 31 BP; 5 A; 8 C; 12 G; 6 T; 0 other;

AAA51910 Length: 31 March 5, 2002 14:18 Type: N Check: 5916
Found using 'seq2-3' (pappu403.key)

1 match found in sequence:

aaa51918 ; T helper cell differentially expressed gene.
(from "mycobacterng.seq")
TOIG of: aaa51918 check: 4217 from: 1 to: 843

1 CCGCGGGTACCAGTAAATCGTCCTGGGGGTGG
17 22

1 match found in sequence:

aaa51916 ; Forward primer to amplify murine 103 gene probe.
(from "mycobacterng.seq")
TOIG of: aaa51916 check: 6990 from: 1 to: 21

ID AAA51916 standard; DNA; 21 BP.

XX AAA51916;

AC AAA51916;

XX 31-OCT-2000 (first entry)

DT Forward primer to amplify murine 103 gene probe.

DE T helper cell; differential expression; 103 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; primer; ss.

OS Mus sp.

XX US6084083-A.

PN 04-JUL-2000.

PD 28-MAR-1997; 9705-0829525.

PF 01-MAR-1996; 9605-0609583.

PR 03-MAR-1995; 9505-0398633.

PR 07-JUN-1995; 9505-0487748.

XX (MILL-) MILLENNIUM PHARM INC.

PA Levinson DA;

XX WPI; 2000-464385/40.

XX New isolated human 200 gene products or polypeptides, useful for
XX treating and diagnosing immune disorders, especially T helper
XX lymphocyte-related disorders

PS Example; Column 93; 107pp; English.

XX Genes which are differentially expressed within and among T helper (TH)
XX cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
XX can be used diagnostically or as targets for therapeutic intervention.
XX The polypeptides are useful for treating and diagnosing of immune
XX disorders, especially T lymphocyte-related disorders. These disorders
XX include chronic inflammatory diseases and disorders (e.g. Crohn's
XX disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
XX Grave's disease), or atopic conditions (e.g. asthma and allergy,
XX including allergic rhinitis or food allergies). Also included are
XX certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
XX HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.

SQ Sequence 21 BP; 3 A; 4 C; 8 G; 6 T; 0 other;

AAA51916 Length: 21 March 5, 2002 14:18 Type: N Check: 6990
Found using 'seq2-3' (pappu403.key)

1 GTAAATCGTCCTGGGGGTCTGG
5 10

2 matches found in sequence:

aaa51918 ; T helper cell differentially expressed gene.
(from "mycobacterng.seq")
TOIG of: aaa51918 check: 4217 from: 1 to: 843

SQ Sequence 903 BP; 242 A; 225 C; 221 G; 215 T; 0 other;

AAA51905 Length: 903 March 5, 2002 14:18 Type: N Check: 40
Found using 'seq2-3' (pappu403.key)

...

608 TCTACATCGAGCAGGGATCTGTGCTGGCTGGCTCTTATCTTCGGCGCTTAA
|-----|
658 663

668 TTTTCAAGTATTCTCATAGCAAGAGAAGATACAGAAATTTAAG

...

837 GTATTATTGCTATGTCAGCAGCAGGAGCAACCCCTCAACACCTTTGGTGTGCGCTTGC
|-----|
887 892

897 AATGCCA

1 match found in sequence:

aaa51907 ; Reverse primer for murine gene 200 fragment amplification.
(from "mycobacterng.seq")
TOIG of: aaa51907 check: 7163 from: 1 to: 39

ID AAA51907 standard; DNA; 39 BP.

XX AC AAA51907;

XX DT 31-OCT-2000 (first entry)

XX DE Reverse primer for murine gene 200 fragment amplification.

XX KW T helper cell; differential expression; 200 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; primer; ss.

OS Mus sp.

XX US6084083-A.

XX PD 04-JUL-2000.

XX PF 28-MAR-1997; 9705-0829525.

XX XX 01-MAR-1996; 9605-0609583.

XX PR 03-MAR-1995; 9505-0398633.

XX PR 07-JUN-1995; 9505-0487748.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX XX Levinson DA;

XX DR WPI; 2000-464385/40.

XX PT New isolated human 200 gene products or polypeptides, useful for

XX PT treating and diagnosing immune disorders, especially T helper

XX PT lymphocyte-related disorders

XX PS Example; Column 88; 107pp; English.

XX CC AAA51906-07 were used to amplify the cDNA encoding the signal sequence
CC and extracellular domain of the murine 200 gene. The amplified fragment
CC was used to construct an IgG1 fusion protein.
CC Genes which are differentially expressed within and among T helper (TH)
CC cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
CC can be used diagnostically or as targets for therapeutic intervention.
CC The polypeptides are useful for treating and diagnosing of immune
CC disorders, especially T lymphocyte-related disorders. These disorders
CC include chronic inflammatory diseases and disorders (e.g. Crohn's

CC disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
CC Grave's disease), or atopic conditions (e.g. asthma and allergy,
CC including allergic rhinitis or food allergies). Also included are
CC certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
CC HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.

XX SQ Sequence 39 BP; 7 A; 10 C; 7 G; 15 T; 0 other;

AAA51907 Length: 39 March 5, 2002 14:18 Type: N Check: 7163
Found using 'seq2-3' (pappu403.key)

1 TTAATTGGATCCCGAGTCTCTGATCGTTCTCTCCAGAGTC
|-----|
23 28

1 match found in sequence:

aaa51908 ; Forward primer for human gene 200 fragment amplification.
(from "mycobacterng.seq")
TOIG of: aaa51908 check: 7788 from: 1 to: 32

ID AAA51908 standard; DNA; 32 BP.

XX AC AAA51908;

XX DT 31-OCT-2000 (first entry)

XX DE Forward primer for human gene 200 fragment amplification.

XX KW T helper cell; differential expression; 200 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; primer; ss.

OS Homo sapiens.

XX US6084083-A.

XX PD 04-JUL-2000.

XX PF 28-MAR-1997; 9705-0829525.

XX PR 01-MAR-1996; 9605-0609583.

XX PR 03-MAR-1995; 9505-0398633.

XX PR 07-JUN-1995; 9505-0487748.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX XX Levinson DA;

XX DR WPI; 2000-464385/40.

XX PT New isolated human 200 gene products or polypeptides, useful for

XX PT treating and diagnosing immune disorders, especially T helper

XX PT lymphocyte-related disorders

XX PS Example; Column 89; 107pp; English.

XX CC AAA51908-09 were used to amplify the cDNA encoding the signal sequence
CC and extracellular domain of the human 200 gene. The amplified fragment
CC was used to construct an IgG1 fusion protein.
CC Genes which are differentially expressed within and among T helper (TH)
CC cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
CC can be used diagnostically or as targets for therapeutic intervention.
CC The polypeptides are useful for treating and diagnosing of immune
CC disorders, especially T lymphocyte-related disorders. These disorders
CC include chronic inflammatory diseases and disorders (e.g. Crohn's
CC disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
CC Grave's disease), or atopic conditions (e.g. asthma and allergy,
CC including allergic rhinitis or food allergies). Also included are
CC certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
CC HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.


```
...
248 CTGATGAAGAAATGTGACATATCAGAAATCCAGCAGATACCAGCTAAAGGGCGATCTCA
    |-----|
    298 303
308 ACAAGGAGATGTCTCTGATCATATAAGAAATGTGACTCTGGATGA
...
553 AATGGAACAAAATTTCCACATGGGCTGATGAATTAAGGACTCTGGAGAAACGATCAGA
    |-----|
    603 608
613 ACTGCTATCCACATTCGAGTGGGAGTCTCTGCTGGGTTGACCCCTGG
...
2201 GAGCTCTTGACAGCAATCTATATAGTCAGCAAAATATTTTGGGAGGACAGTCGTCACCA
    |-----|
    2251 2256
2261 AATTGATTTTCAAGCGGTGGACCTCAGTTTCATCTGCTTACAGC
...
-----
4 matches found in sequence:
aaa51899 ; Murine T helper cell differentially expressed gene 54.
      (from "mycobactereng.seq")
      TOIG of: aaa51899 check: 6440 from: 1 to: 1257

ID   AAA51899 standard; DNA; 1257 BP.
XX
AC   AAA51899;
XX
DT   31-OCT-2000 (first entry)
XX
DE   Murine T helper cell differentially expressed gene 54.
XX
KW   T helper cell; differential expression; 54 gene; immunomodulator;
KW   anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW   thyromimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW   protozoacide; lymphocyte; modulator; gene therapy; cysteine protease;
KW   prepro-sequence; ss.
XX
OS   Mus sp.
XX
FH   Key
FT   CDS
    22..1137
    /tag=a
    /product=Cysteine_protease-like_protein
    /note="The protein product comprises a prepro-protein
    putatively cleaved between amino acid residues
    125 and 136"
XX
PN   US6084083-A.
XX
PD   04-JUL-2000.
XX
PF   28-MAR-1997; 97US-0829525.
XX
PR   01-MAR-1996; 96US-0609583.
PR   03-MAR-1995; 95US-0398633.
PR   07-JUN-1995; 95US-0487748.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Levinson DA;
XX
DR   WPI; 2000-464385/40.
DR   P-PSDB; AAY97057.
```

```
XX
PT   New isolated human 200 gene products or polypeptides, useful for
PT   treating and diagnosing immune disorders, especially T helper
PT   lymphocyte-related disorders
XX
PS   Example; Fig 22A-C; 107pp; English.
XX
CC   The 54 gene appears to encode a novel cysteine protease, inhibition of
CC   which may possibly serve to minimize tissue damage in disorders involving
CC   TH1-like cells.
CC   Genes which are differentially expressed within and among T helper (TH)
CC   cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
CC   can be used diagnostically or as targets for therapeutic intervention.
CC   The polypeptides are useful for treating and diagnosing of immune
CC   disorders, especially T lymphocyte-related disorders. These disorders
CC   include chronic inflammatory diseases and disorders (e.g. Crohn's
CC   disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
CC   Grave's disease), or atopic conditions (e.g. asthma and allergy,
CC   including allergic rhinitis or food allergies). Also included are
CC   certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
CC   HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.
XX
SQ   Sequence 1257 BP; 323 A; 357 C; 323 G; 253 T; 1 other;
    AAA51899 Length: 1257 March 5, 2002 14:18 Type: N Check: 6440 ..
    Found using 'seq2-3' (pappu403.key)
...
488 CTGCGACAACATCCAGGCTCTGTGGCGCATCAAAACACCAGCAGTTTGTGGAGCTCTCTG
    |-----|
    538 543
548 TGCAGGAGCTGTGGACTCGAACGCTGTGGAATGGTTGCAATGGTGGCTTCGTGTGGG
    |-----|
    569 574
608 ACGCATATCTAACTGTC
...
912 CAAGGAGAAAGAGGGCATGCAGACAGGAGACAGTCTTTGTCCCATTCCTCGAAACGTCGCCA
    |-----|
    962 967
    965 970
972 CTCTCCCATACTGGATCCTGAAGAACTCTCTGGGAGCTCACTGGGGC
...
-----
4 matches found in sequence:
aaa51904 ; Human T helper cell differentially expressed gene 200.
      (from "mycobactereng.seq")
      TOIG of: aaa51904 check: 7978 from: 1 to: 2236

ID   AAA51904 standard; DNA; 2236 BP.
XX
AC   AAA51904;
XX
DT   31-OCT-2000 (first entry)
XX
DE   Human T helper cell differentially expressed gene 200.
XX
KW   T helper cell; differential expression; 200 gene; immunomodulator;
KW   anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW   thyromimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW   protozoacide; lymphocyte; modulator; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key
    Location/Qualifiers
```

PR 07-JUN-1995; 95US-0487748.
XX (MILL-) MILLENNIUM PHARM INC.
PA Levinson DA;
XX WPI: 2000-464385/40.
XX P-PSDB; AAY97054.
XX
PT New isolated human 200 gene products or polypeptides, useful for
PT treating and diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders
XX
XX Example; Fig 9A-D; 107pp; English.
XX
CC Genes which are differentially expressed within and among T helper (TH)
CC cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations.
CC The polypeptides are useful for therapeutic intervention.
CC The polypeptides are useful for treating and diagnosing of immune
CC disorders, especially T lymphocyte-related disorders. These disorders
CC include chronic inflammatory diseases and disorders (e.g. Crohn's
CC disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
CC Grave's disease), or atopic conditions (e.g. asthma and allergy,
CC including allergic rhinitis or food allergies). Also included are
CC certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
CC HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.
XX
SQ Sequence 2055 BP; 623 A; 460 C; 415 G; 557 T; 0 other;

AAA51893 Length: 2055 March 5, 2002 14:18 Type: N Check: 1121 ..
Found using 'seq2-3' (pappu403.key)
...
448 TTCCCTGCTGACGCTTAAGAAGTGATGTAACCTGCCACTGTGAGACCATGGCGATGAAC |-----|
498 503
508 AGCATGTGCATTGAAGACGACGCCACCTCGAACACTATTGTTCCGCGTGGTCTACATA |-----|
527 532
568 ATTGTGTTTATAGTCAGCGTCCAGCCACATCGGATCTTTATGCGGTATCCTCTTGTGCAA |-----|
583 588
628 GCGAAGAAGGA
...
719 ACACCTTGGATAAAGACAACACTGGACTTTCTCTCCACCTTGTGCAAGGAGCGTTTCT |-----|
769 774
779 TCACCTACATGAACCTTTTACAGCAGCAGCGCGTTCCTCACTTGCATTTGCCCTGGACGCT |-----|
807 812
839 ATTTAGCAGTCGTCTACCCCTCTGAAGTTTTCCTTCTCCTAAGACGAGAGATTGCGGTTTA |-----|
847 852
899 TTAC
...
1220 CCCCCCTCCACGTGATGGTCTCATCGGCTGGCTTTTAGAGCGGACATGAAGCTCAATG |-----|
1270 1275
1280 ACAAGTCTGGATGGCAGACGCTTACGGGTGTACAGATCAGATAGCCCTGACGAGTCTAA |-----|
1296 1301

1340 ACTGTGTTGCCG
...

3 matches found in sequence:
aaa51898 ; Murine T helper cell differentially expressed gene 200.
(from "mycobactereng.seq")
TOIG of: aaa51898 check: 1752 from: 1 to: 2710

ID AAA51898 standard; DNA; 2710 BP.
XX
XX AAA51898;
XX
XX 31-OCT-2000 (first entry)
XX Murine T helper cell differentially expressed gene 200.
XX
XX T helper cell; differential expression; 200 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; ss.
XX
XX Mus sp.
XX
FH Key Location/Qualifiers
CDS 40..885 /*tag= a
FT sig_peptide 40..99 /*tag= b
FT mat_peptide 100..882 /*tag= c
FT
FT
XX
XX US6084083-A.
XX
XX 04-JUL-2000.
XX
XX 28-MAR-1997; 97US-0829525.
XX
XX 01-MAR-1996; 96US-0609583.
XX 03-MAR-1995; 95US-0398633.
XX 07-JUN-1995; 95US-0487748.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Levinson DA;
XX
XX WPI: 2000-464385/40.
XX P-PSDB; AAY97056.
XX
XX New isolated human 200 gene products or polypeptides, useful for
XX treating and diagnosing immune disorders, especially T helper
XX lymphocyte-related disorders
XX
XX Example; Fig 17A-D; 107pp; English.
XX
XX Genes which are differentially expressed within and among T helper (TH)
XX cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
XX can be used diagnostically or as targets for therapeutic intervention.
XX The polypeptides are useful for treating and diagnosing of immune
XX disorders, especially T lymphocyte-related disorders. These disorders
XX include chronic inflammatory diseases and disorders (e.g. Crohn's
XX disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
XX Grave's disease), or atopic conditions (e.g. asthma and allergy,
XX including allergic rhinitis or food allergies). Also included are
XX certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
XX HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.
XX
XX Sequence 2710 BP; 789 A; 592 C; 644 G; 678 T; 7 other;

AAA51898 Length: 2710 March 5, 2002 14:18 Type: N Check: 1752 ..
Found using 'seq2-3' (pappu403.key)

```
579  ACTCCAGTATTGTCACAGGTCAAATCATCCGGTTCTCCATGTTTCTGGAACGATGCTG
      |-----|
      629 634
639  TTGCCTACTGCACATGGCGGGCAAGAGGTTGCCCTACTGAGGCAGA
...
887  GTGTGGGAGTGACCTCAGACTGGTGACTGTTCCACCATTTCTGTGAGGAACGTTCAAC
      |-----|
      937 942
947  CCAAAGGTCCTCACTTCTGGGAAAGACCGAGTGGAAGAAGGGTGAT
...
1102 CCTGCCACCAGACTGACAGCCAGAGAGGCTTTTCCAGATTCAAGAAGCGTTTCTT
      |-----|
      1152 1157
1162 ACTCGCAGTGCGCTCCCTCGAAATCTGAACATGATCATGTATAA
...
1278 TGGCGACTGTGACACGCTGTGCTTTATTTGTGTGTATCTTTGGGGATCATCGCCATGT
      |-----|
      1328 1333
1338 TTTACTTTGAAGCCTTTTGAAGAGGAGAGCGCGGAGAACCGAGAA
...
-----
3 matches found in sequence:
aaa51892 ; T helper cell differentially expressed gene 103.
(from "mycobacterng.seq")
TOIG of: aaa51892 check: 2381 from: 1 to: 255

ID AAA51892 standard; DNA; 255 BP.
XX
AC AAA51892;
XX
DT 31-OCT-2000 (first entry)
XX
DE T helper cell differentially expressed gene 103.
XX
KW T helper cell; differential expression; 103 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; ST-2; T1; Fit-1; ss.
XX
OS Mus sp.
XX
PN US6084083-A.
XX
PD 04-JUL-2000.
XX
PF 28-MAR-1997; 97US-0829525.
XX
PR 01-MAR-1996; 96US-0609583.
PR 03-MAR-1995; 95US-0398633.
PR 07-JUN-1995; 95US-0487748.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Levinson DA;
XX
DR WPI; 2000-464385/40.
XX
XX New isolated human 200 gene products or polypeptides, useful for
PT treating and diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders
XX
```

```
PS Disclosure; Fig 4A; 107pp; English.
XX
XX The 103 gene, also known as ST-2, T1 or Fit-1, is differentially
XX expressed in TH2-specific subpopulations.
XX Genes which are differentially expressed within and among T helper (TH)
XX cells and TH cell subpopulations, e.g. TH0, TH1 and TH2 subpopulations,
XX can be used diagnostically or as targets for therapeutic intervention.
XX The polypeptides are useful for treating and diagnosing of immune
XX disorders, especially T lymphocyte-related disorders. These disorders
XX include chronic inflammatory diseases and disorders (e.g. Crohn's
XX disease, reactive arthritis, Lyme disease, Hashimoto's thyroiditis or
XX Grave's disease), or atopic conditions (e.g. asthma and allergy,
XX including allergic rhinitis or food allergies). Also included are
XX certain pathogen susceptibilities (e.g. leishmaniasis), and viral (e.g.
XX HIV) or bacterial (e.g. tuberculosis or lepromatous leprosy) infections.
XX
SQ Sequence 255 BP; 82 A; 53 C; 58 G; 62 T; 0 other;
AAA51892 Length: 255 March 5, 2002 14:18 Type: N Check: 2381
Found using 'seq2-3' (pappu403.key)

1 |-----|
3 8
TTAGCGCCATTTGCCATAGAGAGACCTCAGCCATCAATCACTAGACATGATTGACAGA

...

80 TTGGCAATTCGACACTTCCCATGTATTGACAGATTACGGAGGCGACATAATCGTCTCTGG
      |-----|
      130 135

140 GGTCTGGAATAAGGCTTTAATTGTGAGATGCCCCCAAGAGGACCTCGACTTATCCT
      |-----|
      183 188

200 GTGGAATGGTATTACTACATACATAAATGAAGTATTCCT
...
-----
8 matches found in sequence:
aaa51893 ; T helper cell differentially expressed gene 10.
(from "mycobacterng.seq")
TOIG of: aaa51893 check: 1121 from: 1 to: 2055

ID AAA51893 standard; DNA; 2055 BP.
XX
AC AAA51893;
XX
DT 31-OCT-2000 (first entry)
XX
DE T helper cell differentially expressed gene 10.
XX
KW T helper cell; differential expression; 10 gene; immunomodulator;
KW anti-inflammatory; anti-arthritis; antibacterial; immunosuppressive;
KW thymimetic; anti-thyroid; anti-asthmatic; anti-allergic; antiviral;
KW protozoacide; lymphocyte; modulator; gene therapy; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 496..1509
FT /tag= a
FT /product= 7_transmembrane_domain_protein
XX
PN US6084083-A.
XX
PD 04-JUL-2000.
XX
PF 28-MAR-1997; 97US-0829525.
XX
PR 01-MAR-1996; 96US-0609583.
PR 03-MAR-1995; 95US-0398633.
PR
```


CC bone marrow stromal cell cDNA library. TANGO 228 nucleic acids, proteins and modulators can be used to: modulate the proliferation, differentiation and/or function of cells that form the spleen, e.g. to treat (foetal) spleen-associated diseases such as splenic lymphoma and/or splenomegaly, and/or phagocytic disorders such as those inhibiting macrophage engulfment of bacteria and viruses in the bloodstream; to modulate mast cell function and thus to treat immunological disorders and diseases including allergic asthma and atopic dermatitis; to protect the body from antigenic invaders e.g. by modulating the activity of macrophage for treatment of anaphylactic shock or allergic dermatitis; to modulate type I immunological disorders, e.g. anaphylaxis or rhinitis, by modulating the interaction between antigens and mast cell receptors; and to treat tumour necrosis factor-related disorders (e.g. acute myocarditis, myocardial infarction, congestive heart failure), cell disorders (e.g. dermatitis, fibrosis), differentiative and apoptotic disorders, and disorders related to angiogenesis (e.g. tumor formation, metastasis, cancer). Nucleic acids can be used for recombinant production of TANGO 228 polypeptides, in gene therapy and antisense therapy protocols, and as probes and primers for diagnosis.

SQ Sequence 911 BP; 270 A; 207 C; 219 G; 212 T; 3 other;

AAA50444 Length: 911 March 5, 2002 14:18 Type: N Check: 131
Found using 'seq2-3' (pappu403.key)

1 CGGCACAGGCGCGCGCTCGCCATGTGCCTCTCTGCGGTTTCATTCAAGGAATAA
9 14

61 GATG

...

559 ACCAGTCCAAGGTTCTGAGATGCGCCTGCACAGGACGAGTGTATGTCAAGCCTGCA
609 614

619 AGACTCAGACAGACCAACCCAGGAGATACACTATGCTCCACTCCAGT

...

10 matches found in sequence:
aaa50445 ; Mouse TANGO 240 cDNA.
(from "mycobacterng.seq")
TOIG of: aaa50445 check: 3396 from: 1 to: 2426

ID AAA50445 standard; cDNA; 2426 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

31-AUG-2000.

25-FEB-2000; 2000WO-US05035.

26-FEB-1999; 99US-0259387.

(MILL-) MILLENNIUM PHARM INC.

PI Fraser CC;

XX WPI: 2000-533178/48.

DR P-PSDB; AAY95971.

XX Nucleic acids encoding TANGO 228, 240 and 243 pp. which have homology to the rat mast cell Ag-32, the Mycobacterium tuberculosis hypothetical protein Rv0712 and human phospholipase A2-activating protein.

PS Disclosure; Fig 18A-B; 188pp; English.

XX The present sequence is that of mouse cDNA clone jtmMal00b11. It includes a coding region encoding TANGO 240 (see AAY95971), a secreted protein with homology to the Mycobacterium tuberculosis hypothetical protein Rv0712, and 86% overall amino acid sequence identity with human TANGO 240 (see AAA50442). The cDNA clone was isolated from a bone marrow stromal cell cDNA library. TANGO 240 nucleic acids, proteins and modulators can be used to: modulate the proliferation, differentiation and/or function of bone and cartilage cells, e.g. chondrocytes and osteoblasts, and to treat bone and/or cartilage associated diseases and disorders, such as those resulting from injury (bone breakage, cartilage tearing), degeneration (osteoporosis), and degeneration of joints (arthritis, osteoarthritis, bone wearing). Homology to Rv0712 suggests use to treat diseases associated with bacterial infection, e.g. tuberculosis. TANGO 240 can also be used to modulate (e.g. trigger) an immune response to treat an immunological disease such as those associated with the respiratory system, e.g. asthma. TANGO 240 proteins, nucleic acids and modulators can also be used to treat disorders of the cells and tissues in which TANGO 240 is expressed. Nucleic acids can be used for recombinant production of TANGO 240 polypeptides, in gene therapy and antisense therapy protocols, and as probes and primers for diagnosis.

XX Sequence 2426 BP; 640 A; 533 C; 654 G; 599 T; 0 other;

AAA50445 Length: 2426 March 5, 2002 14:18 Type: N Check: 3396
Found using 'seq2-3' (pappu403.key)

...

45 GACTGCGCGAGTCTTCTTGTGCTGTGTTGCGTGGCGGCTCGTGGCGGAGCGATGAGG
95 100

105 CCGAGGCCAGGAGGTGCGGCTCCCTTCGCGGCTCGTGGCGGAGCGGAGCGGAGG
124 129

165 GGGCGGGGCCCATGGCAGCTCGGCGCGCGCGGCGGAGCGCTACTCCCGGAGCGGAGCGGCGG
198 203

225 CGGCGCTGACCTCAGGCCCGCGGCGGCTCGCGGCTCACCAGATGGTCC

...

324 AGGATGGAGAAGCCCTGCCAGGAGAGTCACTGTTGATGGCTTTTACATGAGCGGCGGCGG
374 379

384 AAGTCAGCAATGCGGATTTTGAAGAAGTTTGTGAACTCGACTGGGCTA

...

```
FT      /*tag= a
FT      /note= "a nucleic acid comprising this region
FT      of human TANGO 243 cDNA is also
FT      specifically claimed in Claim 1a; encodes
FT      AAY95968"
FT      CDS      3..2567
FT      /*tag= b
FT      /partial
FT      /note= "alternative coding sequence; encodes
FT      AAY95969"
XX
XX      WO200050443-A2.
XX      31-AUG-2000.
XX
XX      25-FEB-2000; 2000WO-US05035.
XX
XX      26-FEB-1999; 99US-0259387.
XX
XX      (MILL-) MILLENNIUM PHARM INC.
XX
XX      Fraser CC;
XX
XX      WPI; 2000-533178/48.
XX      P-PSDB; AAY95968, AAY95969.
XX
XX      Nucleic acids encoding TANGO 228, 240 and 243 pp. which have homology
XX      to the rat mast cell Ag-32, the Mycobacterium tuberculosis hypothetical
XX      protein RV0712 and human phospholipase A2-activating protein -
XX      Claim 1a; Fig 9A-C; 188pp; English.
XX
XX      The present sequence is that of human cDNA clone jthsa049g04
XX      encoding TANGO 243 (see AAY95968), a protein that includes 6 G-beta
XX      domains and 1 G-beta-like domain, and which has homology to human
XX      phospholipase A2-activating protein. The cDNA was isolated from a
XX      foetal spleen cDNA library. Clone Ept243, which encodes human
XX      TANGO 243, is deposited as ATCC 207116. An alternative translation
XX      of the cDNA clone encodes an 885 amino acid non-full-length protein
XX      (see AAY95968). TANGO 243 nucleic acids, proteins and modulators can
XX      be used to: modulate the proliferation, differentiation and/or
XX      function of cells that form the spleen, e.g. to treat (foetal)
XX      spleen-associated diseases such as splenic lymphoma and/or
XX      splenomegaly, and/or phagocytic disorders such as those inhibiting
XX      macrophage engulfment of bacteria and viruses in the bloodstream;
XX      to treat inflammatory arthropathy, bone and cartilage diseases and
XX      disorders, e.g. rheumatoid arthritis; to modulate chemotaxis,
XX      regulate blood vessel permeability, and promote membrane fusion; to
XX      cause monocytes to secrete interleukin 1 (IL-1) and thereby modulate
XX      T-cell-mediated immunity; to modulate IL-1 and tumour necrosis
XX      factor synthesis and release (e.g. by activating phospholipase A2)
XX      and thereby treat immunological disorders; and to modulate levels
XX      of arachidonic acid and thereby control levels of eicosanoids and
XX      prostaglandins. Nucleic acids can be used for recombinant
XX      production of TANGO 243 polypeptides, in gene therapy and antisense
XX      therapy protocols, and as probes and primers for diagnosis.
XX
XX      Sequence 2811 BP; 852 A; 571 C; 643 G; 745 T; 0 other;
XX
XX      AAA50443 Length: 2811 March 5, 2002 14:18 Type: N Check: 637 ..
XX      Found using 'seq2-3' (pappu403.key)
XX
XX
XX      TCCGGTCTCCGGCGCCCTTACCTCAGGCTTTCTCCGCGCGCGCGGCGGCTCTCC
XX      109 114
XX      |-----|
XX      119 GAGTCCGCCCTCGGGACTGGTCTCGGCACAGTGCCTGGGACCGGCCGACAGACACT
XX      121 126 162 167
XX      179 GGCCATGACGAGCGGCGCAACCAAGGTACCGGCTGAGCTG
```

```
...
1111 TCAAGGCTTTTAAAGAAGAACTGCTCACCACCAACATTGATTCTAAAACTGGCGATTAG
1161 1166 |-----|
1171 GGGACATCAATGCTGAGCAGCTTCTCTGGGAGGAACATCTTAATGA
...
1625 TTCGGGATCTTTAACAACACTACCCACAGCAGATCTTTTACAGGTCGTGCTGTTATGT
1675 1680 |-----|
1685 ACCAGGTTCTGCAAGTATGGGAACACTACCATGGCCGGAGTTGATCCA
...
-----
2 matches found in sequence:
aaa50444 ; Mouse TANGO 228 cDNA.
(from "mycobacterieng.seq")
TOIG of: aaa50444 check: 131 from: 1 to: 911
ID AAA50444 standard; cDNA; 911 BP.
XX
XX AC AAA50444;
XX
XX DT 05-DEC-2000 (first entry)
XX
XX DE Mouse TANGO 228 cDNA.
XX
XX KW TANGO 228; mouse; spleen disorder; immunological disorder;
KW immunomodulator; antiinflammatory; cancer; tumour; metastasis;
KW antitumour; anticancer; antimetastatic; gene therapy; diagnosis;
KW ss.
XX
XX OS Mus musculus.
XX
XX FH Key Location/Qualifiers
XX CDS 27..746
XX FT /*tag= a
XX FT sig_peptide 27..128
XX FT /*tag= b
XX FT mat_peptide 129..743
XX FT /*tag= c
XX
XX WO200050443-A2.
XX
XX 31-AUG-2000.
XX
XX 25-FEB-2000; 2000WO-US05035.
XX
XX 26-FEB-1999; 99US-0259387.
XX
XX PA (MILL-) MILLENNIUM PHARM INC.
XX
XX PI Fraser CC;
XX
XX DR WPI: 2000-533178/48.
XX DR P-PSDB; AAY95970.
XX
XX PT Nucleic acids encoding TANGO 228, 240 and 243 pp. which have homology
XX to the rat mast cell Ag-32, the Mycobacterium tuberculosis hypothetical
XX protein RV0712 and human phospholipase A2-activating protein -
XX
XX PS Disclosure; Fig 14; 188pp; English.
XX
XX CC The present sequence is that of mouse cDNA clone jtmMal07f05
XX encoding TANGO 228 (see AAY95970), a protein with homology to rat
XX MCA-32 (mast cell Ag-32) and 30.6% amino acid sequence identity to
XX human TANGO 228 (see AAY95966). The cDNA clone was isolated from a
```


CC encoding TANGO 228 (see AAY95966), a protein that includes 2 Ig
CC domains and which has homology to rat MCA-32 (masted cell Ag-32), a
CC cell surface antigen that is up-regulated in activated mast cells.
CC The cDNA was isolated from a foetal spleen cDNA library. Clone
CC Ept228, which encodes human TANGO 228, is deposited as ATCC 207116.
CC The gene maps to chromosome 17q23. TANGO 228 nucleic acids,
CC proteins and modulators can be used to: modulate the proliferation,
CC differentiation and/or function of cells that form the spleen, e.g.
CC to treat (foetal) spleen-associated diseases such as splenic
CC lymphoma and/or splenomegaly, and/or phagocytic disorders such as
CC those inhibiting macrophage engulfment of bacteria and viruses in
CC the bloodstream; to modulate mast cell function and thus to treat
CC immunological disorders and diseases including allergic asthma and
CC atopic dermatitis; to protect the body from antigenic invaders e.g.
CC by modulating the activity of macrophage for treatment of
CC anaphylactic shock or allergic dermatitis; to modulate type I
CC immunological disorders, e.g. anaphylaxis or rhinitis, by modulating
CC the interaction between antigens and mast cell receptors; and to
CC treat tumour necrosis factor-related disorders (e.g. acute
CC myocarditis, myocardial infarction, congestive heart failure),
CC T cell disorders (e.g. dermatitis, fibrosis), differentiative and
CC apoptotic disorders, and disorders related to angiogenesis (e.g.
CC tumor formation, metastasis, cancer). Nucleic acids can be used
CC for recombinant production of TANGO 228 polypeptides, in gene
CC therapy and antisense therapy protocols, and as probes and primers
CC for diagnosis.
XX
SQ Sequence 4060 BP; 1162 A; 933 C; 956 G; 1009 T; 0 other;

AAAS0441 Length: 4060 March 5, 2002 14:18 Type: N Check: 9014 ..
Found using 'seq2-3' (pappu403.key)

...

1497 CCTGGGGGATGGGGGACAAACGAGGAGTCACTGGCTAACAGATACAGCGTTTCAG |-----|
1547 1552

1557 TTTGGAAGACAAAAAGTTCTGTGAAAAAGATGGAAGGTGGTGAT

... |-----|
1674 AGTCAGGGGACGATTTCGGCTCACTGCAACCTCTCGCTCCCGGTTTCGGCGATTCTT |-----|
1724 1729

1734 GTGCCTCAGCCTCCAGGTAGCTGGGATGCGAGCGCGCCACCCAGCCGTGCCAGCTAATTT |-----|
1766 1771

1794 TTGTATTTTATAGTGGGATGGGGTTTCA

... |-----|
1985 AAATTTTATGTATGTATACTTACCACAATAATTTAAAAAATATTCTGAAGCGCGCTGCC |-----|
2035 2040

2045 TCATGTGAGTCTTCTTAATAAGAGGGCTGTTCTCTCTTCTCAAGAGCGCTCGCGGGGATT |-----|
2089 2094

2105 GGGAGTTTCTTATCTCAATGTCCACATCCCAAGGCCTA

... |-----|
2395 ACAATCCACATAATTAGCATTTGCCCAACCTTTTCAAGTCTGTGAATGCTGGCGTCCACT |-----|
2445 2450

2455 AGTGCCCTCGTTTCTTTCTAAACCTCATTTCCACATGCGAGGGGAGGTCTTAGGAATGTG

|-----|

2515 GAGCTGTGGCGTTCTTAAGGGTTCTCACTGCTGCTACTATCAACAGAGGGAGTCTCTGTT |-----|
2522 2527
2575 GCC
...
2582 ATCCCTAGGGGTAATTTTGTTCCTGAGGCTGCTTTCTAGGGACTTCTGTGCTCGCTTGTT |-----|
2632 2637
2642 TTATCTGGACCAAGACCTGAAGCAGAGCCTGAAATAAGGCCTTCT
...
2845 TATTGTCCACTTGAGGATTGATAGTAGAGGACATTTATCCACTGGTGTCCAAACGCTCACT |-----|
2895 2900
2905 GGTTGAGGATCACCCAGAAAGCGACACCTCCCCCACTTCTTAGACTAGCATGTGGGTATT |-----|
2925 2930
2965 CCAAGCAGGCTTACCT
...
3922 GGAGGTGGAGGTTCCAGTGAGCCCAAGATCGGGCCACTGTCATTCAGAGCTGGGCCACAGAG |-----|
3972 3977
3982 CAAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
...

10 matches found in sequence:
aaa50442 ; Human TANGO 240 cDNA.
(from "mycobacteriophage")
TOIG of: aaa50442 check: 1254 from: 1 to: 2165

ID AAA50442 standard; cDNA; 2165 BP.
XX
AC AAA50442;
XX
XX 05-DEC-2000 (first entry)
XX
DE Human TANGO 240 cDNA.
XX
KW TANGO 240; human; bone disorder; cartilage disorder; cancer;
KW immunomodulator; antitumour; anticancer; gene therapy; diagnosis;
KW SS.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 2..1126
FT /*tag= a
FT /note= "a nucleic acid comprising this region
FT of human TANGO 240 cDNA is also
FT specifically claimed in Claim 1a"
FT
FT sig_peptide 2..103
FT /*tag= b
FT mat_peptide 104..1123
FT /*tag= c
FT
XX WO200050443-A2.
PN
XX
PD 31-AUG-2000.
XX
XX 25-FEB-2000; 2000WO-US05035.
XX
XX 26-FEB-1999; 99US-0259387.
PR

1 match found in sequence:
aaa16755 ; Human secreted protein clone LL89_3 probe SEQ ID NO:218.
(from "mycobacterng.seq")
TOIG of: aaa16755 check: 842 from: 1 to: 29

ID AAA16755 standard; DNA; 29 BP.
XX
AC AAA16755;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone LL89_3 probe SEQ ID NO:218.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antitumour; antineoplastic; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; hybridisation; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
XX
PR 17-AUG-1998; 98US-0096815.
XX
PR 04-SEP-1998; 98US-0099229.
XX
PR 23-OCT-1998; 98US-0105368.
XX
PR 08-JAN-1999; 99US-0115234.
XX
PR 12-FEB-1999; 99US-0119931.
XX
PR 18-FEB-1999; 99US-0120575.
XX
PR 30-APR-1999; 99US-0132020.
XX
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI; 2000-205979/18.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antinflammatory or tumor inhibition activity
XX
XX Disclosure; Page 631; 641pp; English.

AAA16618 to AAA16697 encode the human secreted proteins given in
AA94988 to AA94980, isolated from human adult brain, adult thyroid,
adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
adult placenta, adult testis, whole embryo, adult cartilage, kidney,
foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
and adult bladder, cDNA libraries. The polynucleotides and proteins are
predicted to have biological activities which would make them suitable
for treating, preventing or ameliorating medical conditions in humans
and animals. The polynucleotides can be used as markers for tissues in
which the protein is preferentially expressed, as molecular weight
markers on Southern gels, and as chromosome markers or tags to identify
chromosomes or to map gene positions. The proteins can be used in the
treatment of immune deficiencies and disorders, such as severe combined
immunodeficiency (SCID), as well as viral, bacterial, fungal and other
infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 29 BP; 7 A; 9 C; 6 G; 6 T; 1 other;

AAA16755 Length: 29 March 5, 2002 14:18 Type: N Check: 842
Found using 'seq2-3' (pappu403.key)

1 CNCTGGTCATAAGACAGTACTCCAGCGCT |----|
24 29

11 matches found in sequence:

aaa50441 ; Human TANGO 228 cDNA.
(from "mycobacterng.seq")

TOIG of: aaa50441 check: 9014 from: 1 to: 4060

ID AAA50441 standard; cDNA; 4060 BP.

XX
AC AAA50441;

XX
DT 05-DEC-2000 (first entry)

XX
DE Human TANGO 228 cDNA.

XX
KW TANGO 228; human; chromosome 17q23; spleen disorder;

XX
KW immunological disorder; immunomodulator; antineoplastic;

XX
KW cancer; tumour; metastasis; antitumour; anticancer; antimetastatic;

XX
KW gene therapy; diagnosis; ss.

XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT CDS 51..1082

FT /*tag= a

FT /note= "a nucleic acid comprising this region

FT of human TANGO 228 cDNA is also

FT specifically claimed in Claim 1a"

FT sig_peptide 51..107

FT /*tag= b

FT mat_peptide 108..1079

FT /*tag= c

XX
PN WO200050443-A2.

XX
PD 31-AUG-2000.

XX
PF 25-FEB-2000; 2000WO-US05035.

XX
PR 26-FEB-1999; 99US-0259387.

XX
XX (MILL-) MILLENNIUM PHARM INC.

XX
PI Fraser CC;

XX
XX WPI; 2000-533178/48.

XX
DR P-PSDB; AA95966.

XX
XX Nucleic acids encoding TANGO 228, 240 and 243 pp. which have homology

XX to the rat mast cell Ag-32, the Mycobacterium tuberculosis hypothetical

XX protein Rv0712 and human phospholipase A2-activating protein -

XX Claim 1a; Fig 3A-H; 188pp; English.

XX
XX The present sequence is that of human cDNA clone jthsa055f08

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX
 SQ Sequence 29 BP; 5 A; 13 C; 3 G; 7 T; 1 other;

AAA16701 Length: 29 March 5, 2002 14:18 Type: N Check: 772
 Found using 'seq2-3' (pappu403.key)

1 GNCATGCTCCTCCTCCTTCAACATCCACG
 4 9
 |-----|

 1 match found in sequence:
 aaal6716 ; Human secreted protein clone rd810_6 probe SEQ ID NO:179.
 (from "mycobacterng.seq")
 TOIG of: aaal6716 check: 2363 from: 1 to: 29

ID AAA16716 standard; DNA: 29 BP.
 XX
 AC AAA16716;
 XX
 DT 16-JUN-2000 (first entry)
 XX
 DE Human secreted protein clone rd810_6 probe SEQ ID NO:179.
 XX

KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; hybridisation; probe; ss.

XX Homo sapiens.
 XX
 XX WO200009552-A1.
 XX
 XX 24-FEB-2000.
 XX
 XX 13-AUG-1999; 99WO-US18298.
 XX
 XX 14-AUG-1998; 98US-0096622.
 XX 17-AUG-1998; 98US-0096815.
 XX 04-SEP-1998; 98US-0099229.
 XX 23-OCT-1998; 98US-0105368.
 XX 08-JAN-1999; 99US-0115234.
 XX 12-FEB-1999; 99US-0119931.
 XX 18-FEB-1999; 99US-0120575.
 XX 30-APR-1999; 99US-0132020.
 XX 11-AUG-1999; 99US-0096622.
 XX (GEMY) GENETICS INST INC.
 XX
 XX Jacobs K, McCoy JW, LaVallie ER, Collins-Racie LA, Evans C;
 PI Werberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX
 XX WPI; 2000-205979/18.

XX New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity
 XX
 PS Disclosure; Page 621; 641pp; English.

XX AAA16618 to AAA16697 encode the human secreted proteins given in

CC AA94898 to AA94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosis, rheumatoid arthritis, autoimmune thyroiditis, insulin
 CC Guillain-Barre syndrome, autoimmune thyroiditis, autoimmune pulmonary inflammation,
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.

XX Sequence 29 BP; 1 A; 8 C; 9 G; 10 T; 1 other;

AAA16716 Length: 29 March 5, 2002 14:18 Type: N Check: 2363
 Found using 'seq2-3' (pappu403.key)

1 TNCCTAGGGCTGCCTGGCTCTGTGCTGCT
 |-----|
 24 29

 1 match found in sequence:

aaal6744 ; Human secreted protein clone cw1292_8 probe SEQ ID NO:207.
 (from "mycobacterng.seq")
 TOIG of: aaal6744 check: 139 from: 1 to: 29

ID AAA16744 standard; DNA: 29 BP.

XX AAA16744;

XX 16-JUN-2000 (first entry)

XX Human secreted protein clone cw1292_8 probe SEQ ID NO:207.

XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; hybridisation; probe; ss.

XX Homo sapiens.

XX WO200009552-A1.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18298.

XX 14-AUG-1998; 98US-0096622.

XX 17-AUG-1998; 98US-0096815.

XX 04-SEP-1998; 98US-0099229.

XX 23-OCT-1998; 98US-0105368.

XX 08-JAN-1999; 99US-0115234.

XX 12-FEB-1999; 99US-0119931.

XX 18-FEB-1999; 99US-0120575.

CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 3580 BP; 734 A; 942 C; 1024 G; 880 T; 0 other;
AAA16697 Length: 3580 March 5, 2002 14:18 Type: N Check: 7693 ..
Found using 'seq2-3' (pappu403.key)

...
325 GGCATTCCCGGCGCAGAGTCTGGGGGCAAGCGCGGATGCGCAGATGGGGCCCGCTG
375 380 |-----|
385 GCTGGAGTAAGGGGGCTTGAGTGATGGGAGGGTCCCTGGGGCCGGGGGAGGCA
445 |-----|
CAACGACGGAAAGACGAGAAGAACTGGCCCAACAGTATGAGCCATCTCTACGGGAGTGT
446 451
505 GGCACGGCGGCTTCAGTGGACACTGATTTTGTGCTGTGGCTGGCGGTGATGGCTGAC
549 554 |-----|
565 GGTGTGAGGTCTTTGTGTGGGCTTCGTCTGCCAGCGCTGAGAAAGACATGTGCCTG
601 606 |-----|
625 TCCGACTCCAACAAGGCATGCTAGGCCTCATGCTACTCTGGGCATGATGTTGGGAGCC
655 660 |-----|
685 TTCTCTGGGAGGTCTGCTGACCGGCTGGGTGCGAGGAGTGTCTGCTCATCTCGCTC
745 TCAGTCAACAGCGTCTTCGCCTTCTTCTCATCTTTTGTCCAGGGTTTACGGCACTTTCCCTC
754 759 |-----|
805 TTCTG
...
1171 ATCGAGCCAAAGGACATCTGTAGCGAGTGTCTCAGTAACCCACATTAAGACGATTTCAT
1221 1226 |-----|
1231 CAGGAGGATGAATTGATTGATGCCAGTCCAGTCGGACACAGGACCTGTGTACCGCGCTGGGG
1280 1285 |-----|
1291 GTCCGGGCTTGAGCCTAGGGGGCAGGTTTGGGGGAATTTTCTC
...
1931 CCTGTCTCTCTCTTCTTCTCTCTTTTGGGAACAGTGTGCGGCATGTCGCTCTCG
1981 1986 |-----|
1991 TCTGCCTTTTGGCGGGGTGACGATTTGCATCTCGGAATCGCTGGA
...

1 match found in sequence:
aaa16701 : Human secreted protein clone pj193_5 probe SEQ ID NO:164.
(from "mycobactering.seq")
TOIG of: aaa16701 check: 772 from: 1 to: 29

ID XX
XX AAA16701 standard; DNA; 29 BP.
AC AAA16701;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone pj193_5 probe SEQ ID NO:164.
XX
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; hybridisation; probe; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY) GENETICS INST INC.
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity -
PS Disclosure; Page 617; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
XX AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX which the protein is preferentially expressed, as molecular weight
XX markers on Southern gels, and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX candidiasis. The proteins can be used to treat autoimmune disorders such
XX as connective tissue disease, multiple sclerosis, systemic lupus
XX erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
XX Guillain-Barre syndrome, autoimmune thyroiditis, insulin independant


```
61  AAGCAGAATGTCGTATCCAGGTTGTGGATAAATTCAAAGGCTTTTCAATTGCACAGA 119
    71 76
---|
121 CGTCTGTGAGACCAACGACTCACGTGCTTTCGGGAAGCCACTTCGCACCCCTGAA 124
    124
...
317 TGCCACTTGTCTCAGGGCGGTACCGGGAACCCCTTTTCGCAGCAGCCAGCATGTTT 367 372
    |-----|
377 GTCTCGCTGCCAGACGCCCCAGTGGCCAGCTCTGTGAAGTCTAGTCCACCTGTGCGGA
437 GGCCGGGTACGCAAGTCCCGCCAGCCAGCATCTCATCGGCGCCTACAGCGGAAG 470 475
    |-----|
497 AAGNAAGCCACAGTCAGTATCTGTCTGAGAAATGGTCTTAGTAAGATCCAGGCACA
557 CAGACGCTGTGGTGTGTGCAGATCTGTGGACAGGTTTCAGGGAGGGCGGCTCAGGCT 559 564
    |-----|
617 CACACCCCTTCCACGACGCTGGGCACCTGGGTGATGTCCTCAGCCTCCAGCATCTGCC
    |-----|
677 CTGGCAGCTGTGTGTGTGCAGATCTGTGGACAGGTTTCAGGGAGGGCGGCTCAGGCT 682 687
    |-----|
737 T
    682 687
...
864 TAGACCTCTCCCTCCCGAGTCCCTTGTCTGTCAGGAGTCCCTGGCAGCGCCGGC 914 919
    |-----|
924 ACTGGGGCCCAAGCCCGCTCCTGTGATCTCTCTCCCAGGTACATCTCATGATCACTC
    |-----|
984 CGTCTGCTCATGTGCTCAAGGGGTGTTAAAGAGCTCAACAGACTCCATCTTTTATTGA 1015 1020 1027
    |-----|
1044 CAAAGTGAGCACAGTGTGACCGTAATGTCCCACTCTGGCGTTCATGTGAGCTGGCCAGGC 1080 1085
    |-----|
1104 GCGTGTCCGATTCTGGGGAGGAAGAGTGGTAGGAGCTGAGCTGAGATCGGA 1106
    1106
...
1232 AGGAAGTGTGAAGCAAGACGCGCTGGGGAATGCGGGAAGCAGGAGCAGCGTCTGTG 1282 1287
    |-----|
1292 CTAGAAATTACCTGCCCTGTGTGGAGTCATATGTGGCGGCAAG
...
1454 CGCATTGTCTCTCAGTCTTGTGAGGCTGGAAGTCTAAGATGGGTATCGGACGCGTGTGTT 1504 1509
    |-----|
1514 TCCCCTCAGGCGCTCTCTCTGGGCTTGCAGACAGCTGCCTTCTTCC
...
-----
9 matches found in sequence:
```

```
aaal6697 ; Human secreted protein clone asl80_1 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6697 check: 7693 from: 1 to: 3580
ID AAA16697 standard; cDNA; 3580 BP.
XX
XX AAA16697;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone asl80_1 nucleotide sequence SEQ ID NO:159.
XX
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
XX P-PSDB; AAY94977.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity
XX
XX Claim 168; Page 613-614; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
XX AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX which the protein is preferentially expressed, as molecular weight
XX markers on Southern gels, and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
```

PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94974.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 162; Page 608; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1724 BP; 574 A; 329 C; 354 G; 467 T; 0 other;

AAA16694 Length: 1724 March 5, 2002 14:18 Type: N Check: 8813
Found using 'seq2-3' (pappu403.key)

...
25 CTCACGCTCGGCGAGCAGCTTCTCTCGTCCGCTGGGACCCCGTCGTGCGCCAGG
I-----I
75 80
85 CCGAGCAGCATGCCCCCTAAAAAGGGAGGTGATGGAATTAACAC

...

15 matches found in sequence:
aaal6696 : Human secreted protein clone ys20_1 nucleotide sequence SEQ ID NO:1
(from "mycobactering.seq")
TOIG of: aaal6696 check: 3672 from: 1 to: 2229

ID AAA16696 standard; cDNA; 2229 BP.
XX
AC AAA16696;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone ys20_1 nucleotide sequence SEQ ID NO:157.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;

KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94976.
XX
CC New polynucleotides encoding secreted proteins, which may have e.g.
CC nutritional, chemokine, immune stimulating or suppressing,
CC hematopoiesis regulating, tissue growth, activin/inhibin
CC antiinflammatory or tumor inhibition activity
XX
PS Claim 166; Page 611-612; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2229 BP; 496 A; 624 C; 535 G; 574 T; 0 other;

AAA16696 Length: 2229 March 5, 2002 14:18 Type: N Check: 3672
Found using 'seq2-3' (pappu403.key)

1
I-----I
4 9
GAAGACCTTGGACACTTGGGCCAACAAAGACATTAGTCATGACAAAGCATGCCACTCTGA
-----I
4
-----I
-----I

1 match found in sequence:
aaal6693 ; Human secreted protein clone pe246_1 nucleotide sequence SEQ ID NO: 1330
(from "mycobacterng.seq")
TOIG of: aaal6693 check: 5864 from: 1 to: 1330

ID AAAL6693 standard; cDNA; 1330 BP.
XX
AC AAAL6693;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone pe246_1 nucleotide sequence SEQ ID NO:151.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Rechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94973.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antinflammatory or tumor inhibition activity
XX
PS Claim 160; Page 606-607; 64lpp; English.
XX
CC AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1330 BP; 364 A; 285 C; 309 G; 372 T; 0 other;
AAAL6693 Length: 1330 March 5, 2002 14:18 Type: N Check: 5864 ..
Found using 'seq2-3' (pappu403.key)
...
733 ACCAGCTCTGTACTGGCTGCATCAGGGATCTTAATCAACACATTTAGCTTGGCGTTTAT |----|
783 788
793 TCATTCCATCACCCCTTACTGTAACTACTATGGCAACTCAATAATT
...

1 match found in sequence:
aaal6694 ; Human secreted protein clone qf122_3 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6694 check: 8813 from: 1 to: 1724

ID AAAL6694 standard; cDNA; 1724 BP.
XX
AC AAAL6694;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone qf122_3 nucleotide sequence SEQ ID NO:153.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94969.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX Claim 152; Page 599-600; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC foetal placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 1148 BP; 325 A; 283 C; 247 G; 293 T; 0 other;

AAA16689 Length: 1148 March 5, 2002 14:18 Type: N Check: 5746
Found using 'seq2-3' (pappu403.key)

104 GCAGCGGGAGCCGCGAGCCCGCGCGCGCGCGCGCGCGCGCTTCGAGGGCGGCCAG
154 159
164 GCGGCGCATGTGTAAGGTGAGCTTCAACTCCGCTCTGCCCGCAAGAGGAGGCAAGAGG
183 188
224 ACGAGCCCAAGAGCGGAGGAGCGCTCATCATCCCCCGACGCGCTCGCGGTGGACT
246 251
284 GCAAGACCCAGATGATGTGTACCATGTGGGCCAAG

1 match found in sequence:
aaa16690 ; Human secreted protein clone dm365_3 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaa16690 check: 9007 from: 1 to: 1353
ID AAA16690 standard; cDNA; 1353 BP.
XX
AC AAA16690;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone dm365_3 nucleotide sequence SEQ ID NO:145.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX
PA
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94970.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 154; Page 601; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the

KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX
XX 17-AUG-1998; 98US-0096815.
XX
XX 04-SEP-1998; 98US-0099229.
XX
XX 23-OCT-1998; 98US-0105368.
XX
XX 08-JAN-1999; 99US-0115234.
XX
XX 12-FEB-1999; 99US-0119931.
XX
XX 18-FEB-1999; 99US-0120575.
XX
XX 30-APR-1999; 99US-0132020.
XX
XX 11-AUG-1999; 99US-0096622.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Fachtel K;
XX
XX WPI: 2000-205979/18.
XX
XX P-PSDB; AAY94967.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity
XX
XX Claim 148; Page 597-598; 641pp; English.

XX AAA16618 to AAA16697 encode the human secreted proteins given in
XX AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX which the protein is preferentially expressed, as molecular weight
XX markers on Southern gels, and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX candidiasis. The proteins can be used to treat autoimmune disorders such
XX as connective tissue disease, multiple sclerosis, systemic lupus
XX erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
XX Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
XX diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
XX autoimmune inflammatory eye disease. The proteins can also be used to
XX treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
XX probes for the human secreted proteins from the present invention.
XX
XX Sequence 1785 BP; 603 A; 373 C; 358 G; 451 T; 0 other;
XX
XX AAA16687 Length: 1785 March 5, 2002 14:18 Type: N Check: 7876
XX Found using 'seq2-3' (pappu403.key)

...
699 GGTGGGCAGATTAGGCGTGGCTTTGTTGCTTTCATGTTACCATCATAGCGTTTGTG
749 754
759 CCACCTTACAAGGAGGAGGAAAAAATCAATTCCTGTGTCGCCAACCCAGACA
...
844 GCACAGATCAGACTCCTCTCTCCAAAGAGCTGTGTGGCTTCACCTGCAAGGGCAGCTCT
894 899
904 TCCACAAGCAGAGCGCCAGGACAAAAAGAGGACCTGTGTGCGCAAGAGCGTTTCCTTG
942 947
964 GTTTCCTCAGCGCGCCAAAGCGGAGTGGCGCCCTCCACCACAGGGCCCCCTAATGGCGG
975 980
1024 -|
CTTTGTCTCGGGGCGAGGAGGACCTCATTAGAAGCGCGCTGGTCTAAAGGGAATGCA
1025 1059 1064
1084 TTTCCAGAACAGGAGGTTTCATCATCTCCTAGCGTTAGCGACAGAGTGGTGCACAGAAAGCTC
1113 1118
1119 1124
1144 TGTGGAGCTATTTTCCAGCGCTTCAGTTACATCAAGGATGGTGTGCTGCACTGGCGGAAAG
1160 1165
1204 GCCTCAGGAGG
...

4 matches found in sequence:
aaal6689 ; Human secreted protein clone da529_3 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6689 check: 5746 from: 1 to: 1148
ID AAA16689 standard; cDNA; 1148 BP.
XX
XX AAA16689;
XX
XX 16-JUN-2000 (first entry)
XX Human secreted protein clone da529_3 nucleotide sequence SEQ ID NO:143.
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
XX antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
XX antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
XX connective tissue disease; multiple sclerosis; erythematosis;
XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
XX insulin dependent diabetes mellitus; graft-versus-host-disease;
XX autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.

CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 839 BP; 285 A; 161 C; 180 G; 213 T; 0 other;

AAA16685 Length: 839 March 5, 2002 14:18 Type: N Check: 7077 ..
Found using 'seq2-3' (pappu403.key)

40 GCTTTGGCCCTTCTTTGTAGGAGAGTTTCATCCGCCCTGAAATCTTCCGATCGTTAAATA
90 95 |-----|

100 ACTCCTCAGGTCCTGCCCTGCACAGGGTTTTTCTTAGTTTGTTCG

567 GTTCAAGCAGTGGTTATACCTTGGCACCTAAATAAAGATACATAAGGAACGATTCTT
617 622 |-----|

627 TGCCCTTGCCTTATGTGGACCATTTTCTCAGACTGGAACTGAGAAA

1 match found in sequence:
aaal6686 ; Human secreted protein clone rk80_3 nucleotide sequence SEQ ID NO:1
(from "mycobacterng.seq")
TOIG of: aaal6686 check: 7307 from: 1 to: 1067

ID AAA16686 standard; cDNA; 1067 BP.

XX AC AAA16686;

XX DT 16-JUN-2000 (first entry)

XX DE Human secreted protein clone rk80_3 nucleotide sequence SEQ ID NO:137.

XX KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.

XX OS Homo sapiens.

XX PN WO200009552-A1.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18298.

XX PR 14-AUG-1998; 98US-0096622.

XX PR 17-AUG-1998; 98US-0096815.

XX PR 04-SEP-1998; 98US-0099229.

XX PR 23-OCT-1998; 98US-0105368.

XX PR 08-JAN-1999; 98US-0115234.

XX PR 12-FEB-1999; 99US-0119931.

XX PR 18-FEB-1999; 99US-0120575.

XX PR 30-APR-1999; 99US-0132020.

PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB: AAY94966.

XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX

PS Claim 146; Page 596-597; 641pp; English.

XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC foetal placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.

SQ Sequence 1067 BP; 293 A; 255 C; 281 G; 238 T; 0 other;

AAA16686 Length: 1067 March 5, 2002 14:18 Type: N Check: 7307 ..
Found using 'seq2-3' (pappu403.key)

...

596 CCAC TGGGAGACTCAGCAGATTCCAAAGCCTCAGTCCAGCCAGCCATGCGGTCCTCT
648 653 |-----|

658 TCTCCGCTTCAAAATCTTCGACGCTCCAGGCTTTGTGGCTGTA

...

9 matches found in sequence:
aaal6687 ; Human secreted protein clone au36_42 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6687 check: 7876 from: 1 to: 1785

ID AAA16687 standard; cDNA; 1785 BP.

XX AC AAA16687;

XX DT 16-JUN-2000 (first entry)

XX DE Human secreted protein clone au36_42 nucleotide sequence SEQ ID NO:139.

XX KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;

PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fecthel K;
 XX
 DR WPI: 2000-205979/18.
 DR P-PSDB: AAY94964.
 XX
 PT New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity
 XX
 PS Claim 142; Page 594; 64lpp: English.
 XX
 CC AAA16618 to AAA16697 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX
 SQ Sequence 1564 BP; 386 A; 427 C; 410 G; 341 T; 0 other;

AAA16684 Length: 1564 March 5, 2002 14:18 Type: N Check: 1191 ..
 Found using 'seq2-3' (pappu403.key)

...
 421 TGGGCCCTTGGCCGTGAGAGGGATTGACAGCCGAGATTTCATAGCGCGATGGCT
 |-----|
 471 476
 481 CCCACTGCCAGGCATCAGCCTTCTGCTAGTCAATCACTGCCCTGG
 ...
 702 AGACCACCTGGTTTATGTGACAGGACTTGCATTCTCCTGGAACATGAGGAACCCGGAG
 |-----|
 752 757
 762 GAAGAGCAAGTGGCAGGAGGAAGAACTTGTGCCAAATATGGGTGAG
 ...

 2 matches found in sequence:

aaal6685 : Human secreted protein clone rj214_14 nucleotide sequence SEQ ID NO
 (from "mycobacterng.seq")
 TOIG of: aaal6685 check: 7077 from: 1 to: 839

ID AAA16685 standard; cDNA; 839 BP.

XX AC AAA16685;

XX 16-JUN-2000 (first entry)

DE Human secreted protein clone rj214_14 nucleotide sequence SEQ ID NO:135.

KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosus;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.

XX OS Homo sapiens.

XX PN W0200009552-A1.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18298.

XX PR 14-AUG-1998; 98US-0096622.

XX PR 17-AUG-1998; 98US-0096815.

XX PR 04-SEP-1998; 98US-0099229.

XX PR 23-OCT-1998; 98US-0105368.

XX PR 08-JAN-1999; 99US-0115234.

XX PR 12-FEB-1999; 99US-0119931.

XX PR 18-FEB-1999; 99US-0120575.

XX PR 30-APR-1999; 99US-0132020.

XX PR 11-AUG-1999; 99US-0096622.

XX (GEMY) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

XX PI Wong GG, Clark HF, Fecthel K;

XX DR WPI: 2000-205979/18.

XX DR P-PSDB: AAY94965.

XX PT New polynucleotides encoding secreted proteins, which may have e.g.

XX PT nutritional, chemokine, immune stimulating or suppressing,

XX PT hematopoiesis regulating, tissue growth, activin/inhibin

XX PT antiinflammatory or tumor inhibition activity

XX PS Claim 144; Page 595; 64lpp: English.

CC AAA16618 to AAA16697 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

...
757 GACATTGGAGCACTAAATGAACCTTGAATGTGTCTGTGGCGAGCAGGATGGTCGCTCTTA 807 812 |-----|
817 CTTTGTGATGAGATCGGGGATGAATGCTCGCTTTAAAAATGCTGCTTTGGATTCTGTTG |-----|
877 CTGGAGAGCTCTCTTTGTTTTGCGGCTGGAACGTTTACAGGGGACGTTTGCAGAGAAG 907 912 919 924 |-----|
937 ATCTGTTCTCGCAATGAGATAGAAGGGACCTACACGTAGACTGTGAAAAAAGGGGCTC |-----|
997 ARAAGTCTGCAGGCTTTCACCTGCCCCGACTTCCAGTTTTTACCATTATTTCTGTCATGGC 1007 1012 |-----|
1057 AATTCC
...
1094 ACTTTTATAATGCGGTTAGTTTGACATGGAACAAATGGCTTGCATGAATCGTTCCGG 1144 1149 |-----|
1154 GGGCTTTTCTGGGCTGCAGCTGGTGAAAGGCTGCACATCAACAACAAGATCAAGT |-----|
1214 CTTTTCGAAAGCAGACTTTTCTGGGCTGGACGATCTGGAATATCTCCAGGCTGAATTTA 1243 1248 |-----|
1274 ATTTATTACGAGATATAGACCGGG
...
1374 GTTCCAGTATGTGCCCATCACCCACCTCGACCTCCGGGGTAACAGGCTGAAACGCTGCCC 1424 1429 |-----|
1434 TATGAGGAGTCTGGAGCAAAATCCCTGGTATTGCGGAGATCCGTCT
...
1618 ACCGAACAGACTTGTCTCTCTTTGAAAACCGAGTGGATTCTAGTCTCCCGGCGCCCTT 1668 1673 |-----|
1678 GCCCAAGAGAGACCTTTGCTCTCGGACCCCTGCCTTCA
...
1771 GGAGGTACAAGATCCCAGGCAACTGSCAGATCAAAATCAGACCCACAGCGGATAGCG 1821 1826 |-----|
- | 1827
1831 ACGGGTAGCTCCAGGAACAAACCTTAGCTAACAGTTTACCCTGCCCTGGGG 1832
...
2033 ACTTTGTGGATTACAAGAACCTTCTCTGTTGGATCTGGCAACAATAACATCGCTACTG 2083 2088 |-----|
2093 TAGAGAACAACACTTTTCAAGAACCTTTTGGACCTCAGGTGGCTATA
...

2171 GGGAGAAATTCGGGGGCTGCAAAACCTAGAGTACCTGAACGTGGAGTACAAACGCTATCC 2221 2226 |-----|
2231 AGCTCATCTCCCGGGCACCTTTCAATGCCATGCCAAACTGAGAT
...
2436 AAACCCCTGGGAGTGTCTCTGCACAAATGTGCGCTTTCAAGCAGTGGGAGAACGCTTGGG 2486 2491 |-----|
2496 TTCCGAAGTCTGATGAGCGACCTCAAGTGTGACACGCCGCTGAACCTCTTTAGAAAGGA 2512 2517 2529 2534 |-----|
2556 TTTTCATGCTCTCTCTCCAATGACGAGATCT
...
2927 ACTCGCTCTCAGACTAAGACCCCCAACCCCAATAGGGAGGCGACAGAGGAGGCATACAT 2977 2982 |-----|
2987 CCTTCCCCACCGCAGGCACCCCGGGGCTGAGGGGCGGTGTACCCA
...
4197 AACGGTTGGGTAGATCCCTTCTCTGAGTACATTAAAGGCTCTTCTCTAATCGCCTTAC 4247 4252 |-----|
4257 TTAGCTTTTACCCTTTAAGTAGCTCTCTTCCCTCGCCCCCACC
...

2 matches found in sequence:
aaal6684 ; Human secreted protein clone qy442_2 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6684 check: 1191 from: 1 to: 1564

ID AAA16684 standard; cDNA; 1564 BP.
XX
AC AAA16684;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone qy442_2 nucleotide sequence SEQ ID NO:133.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.

CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 938 BP; 345 A; 154 C; 197 G; 242 T; 0 other;

AAA16682 Length: 938 March 5, 2002 14:18 Type: N Check: 2860 ..
Found using 'seq2-3' (pappu403.key)

484 AAACTAAGGCCAGATACCAAGATGGAACAGAATTGAAGAACACGAGGATTTAA
534 539
544 GGATTTTACCAGTTTACTTTTAAATTTTGCACAAAGATCCAGGACAA
643 AAGATTTAAATCTTAGACTTATGGAATAAATTTTGTGGACATCATACGATCAAT
693 698
703 ACCAAAGACACTGGGAATCTCTTTAGACTTCAGTACGATGATGAT

21 matches found in sequence:
aaal6683 ; Human secreted protein clone nf56_3 nucleotide sequence SEQ ID NO:1
(from "mycobacterng.seq")
TOIG of: aaal6683 check: 8559 from: 1 to: 5170

ID AAA16683 standard; cDNA; 5170 BP.
AC AAA16683;
XX
DF 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone nf56_3 nucleotide sequence SEQ ID NO:131.
XX
KW Human: secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.

OS Homo sapiens.

XX WO200009552-A1.

PN 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18298.

XX 14-AUG-1998; 98US-0096622.

PR 17-AUG-1998; 98US-0096815.

PR 04-SEP-1998; 98US-0099229.

PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
PA Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fecthel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94963.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 140; Page 590-592; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 5170 BP; 1372 A; 1257 C; 1174 G; 1367 T; 0 other;

AAA16683 Length: 5170 March 5, 2002 14:18 Type: N Check: 8559 ..
Found using 'seq2-3' (pappu403.key)

116 TAAGAGCGGAGGGAGGAAAGGATGGACACACCAACCAACGACGCGATTCGG
166 171
176 GAAATTTTCAGCGCCATTCGGTCGCGAGCGTGAGTCCTTCGGTCGGCGGTGATTCAGC
186 191
236 ACCGGG
616 GCTTGCTTGCACGTGCTCGATAGTTTGTGGATATAATTATTGACTGGCGCTCTGGG
666 671
676 CTATTGCACTGCGGGGGTTAGGAGGAGGAATCCACCCACC

CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAL16698 to AAL16774 represent
CC probes for the human secreted proteins from the present invention..
XX

SQ Sequence 2673 BP; 830 A; 480 C; 670 G; 693 T; 0 other;

AAL16681. Length: 2673 March 5, 2002 14:18 Type: N Check: 3483 ..
Found using 'seq2-3' (pappu403.key)

```
1  AGGTCTAGAAATCAATCGGAGCGCTCGCGCTGATTCTTCACGGGCCGCGTCCGGCCCCC
    20 25
61  CGCTCTGCCCTGCAT
...
854  CTAAGAAAAACATTTTGGAGAATTCAGTAAAGTCACAGAGGCTTTGGTGGACGTATTTC
    |-----|
    904 909
914  TCTATCATCAACCGGATGACAAAAGAAGAAATCGGGGGTCTGCTTCTTGAATATGAGG
...
974  ATCACAAGTCAGCAGCACAGACCGCGCGCTGATGATGGGAAAAGTAAAGTGTGGG
    |-----|
    998 1003
1034  GAAATGTAGTTCAGTTGAA
...
1284  GGAAGAAATGGAATAGTCTTACCCAGCCACCACACAGAAAGGAAGCGCCCAAGC
    |-----|
    1334 1339
1344  TGCTAGACAGGCTCCAGAGCACTGCGTATGAAGATTATTACTAC
...
1782  CAATGTAGGAGGCAAGAGAAAGCAGATGGGTACAACCCAGCCTGATTCCAAGCGTGTCA
    |-----|
    1832 1837
    1835 1840
1842  GACCAACAACCAAGAACTGGGGTTCCCAACCCATCGCTCAGCAGCCGCTTCAGCAAGG
    |-----|
    1876 1881
1902  TGGTGACTATTCTGGTAAGTATGGTTACAA
...
2176  ACATTATTGGGCTTCAGAGTTCCTTATTCTGGGGATTACAATGCTTTTATCGTTTCAG
    |-----|
    2226 2231
2236  GCTTCATTTTAGCTTCAAAAAGAGCTGGGCACACTGTTAAATCATG
...
2503  TATACTTCACACCTTTTGAATCTGCCCTACAAAATTTGTTGGCTTAACGTCAAAA
    |-----|
    2553 2558
2563  GCGGTGACAATTTGTTCTTTTGATGTGATTGATTTCCAAATTTCTTG
```

...

2 matches found in sequence:
aal16682 ; Human secreted protein clone mt124_3 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aal16682 check: 2860 from: 1 to: 938

ID AAL16682 standard; cDNA; 938 BP.
XX
XX AC AAL16682;
XX
XX DT 16-JUN-2000 (first entry)
XX
XX DE Human secreted protein clone mt124_3 nucleotide sequence SEQ ID NO:129.
XX
KW Human: secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX PN WO200009552-A1.
XX
XX PD 24-FEB-2000.
XX
XX PF 13-AUG-1999; 99WO-US18298.
XX
XX PR 14-AUG-1998; 98US-0096622.
XX PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94962.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
XX PS Claim 138; Page 589; 641pp; English.
XX
XX CC AAL16618 to AAL16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify

PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94959.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
PS Claim 132; Page 581-582; 641pp; English.
XX
CC AA16618 to AA16697 encode the human secreted proteins given in
CC AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AA16698 to AA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2509 BP; 748 A; 416 C; 507 G; 838 T; 0 other;

AA16679 Length: 2509 March 5, 2002 14:18 Type: N Check: 6154
Found using 'seq2-3' (pappu403.key)

1 GAGCTGCATCGGGAGCGCATGCGGGGGATGCGCTGCGCGGGCCTGGAGCAGATG
33 38
|-----|
61 TCCTGGTTCTACTACCACTACTGCTGCTCAGCGGCGCTCTACATGCTGGAGCCTGGGAG
93 98
|-----|
121 CGGACGGGTTCAGCTGCTTTCCATTGTGGGATGGCACTATACACAGGATACGCTCTCA
|-----|

181 TGCCCCAGCACATCATGCGGATATTGCACACTTTTGAATCGTACATGACCAAGATGCG
197 202
241 ACCAGGATCAGA
...
1436 GTTGCTGGGTTGTTTTTTTGTGTTTTTTTTTTTGTGAGACGGAGTCTTGCCTGTCGCCGAGG
1486 1491
|-----|
1496 ATGAGTGCAACAGCATGATCTCGGTTTCACTGCAACCTCCATCTCC
...

5 matches found in sequence:
aaal6680 ; Human secreted protein clone mll17_1 nucleotide sequence SEQ ID NO:
(from "mycobactering.seq")
TOIG of: aaal6680 check: 8212 from: 1 to: 2672
ID AA16680 standard: cDNA; 2672 BP.
XX
XX AC AA16680;
XX
XX DT 16-JUN-2000 (first entry)
XX
XX DE Human secreted protein clone mll17_1 nucleotide sequence SEQ ID NO:125.
XX
XX KW Human: secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200009552-A1.
XX
XX PD 24-FEB-2000.
XX
XX PF 13-AUG-1999; 99WO-US18298.
XX
XX PR 14-AUG-1998; 98US-0096622.
XX PR 17-AUG-1998; 98US-0096815.
XX PR 04-SEP-1998; 98US-0099229.
XX PR 23-OCT-1998; 98US-0105368.
XX PR 08-JAN-1999; 99US-0115234.
XX PR 12-FEB-1999; 99US-0119931.
XX PR 18-FEB-1999; 99US-0120575.
XX PR 30-APR-1999; 99US-0132020.
XX PR 11-AUG-1999; 99US-0096622.
XX
XX PA (GEMY) GENETICS INST INC.
XX
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94960.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
XX PS Claim 134; Page 583; 641pp; English.

PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fecthel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94957.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 128; Page 579-580; 641pp; English.
XX
CC AAL16618 to AAL16697 encode the human secreted proteins given in
CC AAY94988 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependant
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAL16698 to AAL16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 4086 BP; 1031 A; 942 C; 1014 G; 1099 T; 0 other;

AAAL16677 Length: 4086 March 5, 2002 14:18 Type: N Check: 4606
Found using 'seq2-3' (pappu403.key)

298 GCAGTTTCCACCCCTCTTCACCTGGAGAACTTGGGCTCCATTAGTGCATCTGTTGGAG
348 353
|-----|

358 TAATTAGCCCATCTTTTACATTTCTTGGCACAAAATCTCGAAGAGC

571 TGTATCCATCTACTGGATAATAGAGTGGGGTGTGGTCTGGTGGTGGGACGATGGTG
621 626
|-----|

631 GTATTGGAGAATACATCCGGCAGTGGTAGCGGCACTATTTCAGAAATCAAAAGACATAGATC
680 685
691 GAGTGTCTCCATGTGATATTGGGATCATGTGTGCTAGCCAGCAACCCCTAGGACACAG
743 748
|-----|
751 GGAAGAATGGAGACTGAGTCACATCAATGACAGCACCTCTCGGTGAT
...
806 GTCAACATATCAGTGTGAGTTTCTGTGGACAACTGGGCCCGCATGCTATGGCGGCCCATC
856 861
|-----|
866 AATTGCTTCCACAAGTATTGGCCAGTCTTGATTCTTAGAGGAATT
...
1027 AGTGTGGAGATGAGGCATCATCTTGTGGGACCTTATTATTCTTGGCCCGGTGCTGAAG
1077 1082
|-----|
1087 GTCCCTCTACAGCGGGCGGGGAGTTTCCCGCGCGGAAGACTTT
...
1251 CCGGTCGGTGGATCTACCTGGTGGTGGTCAATGTGAGGCAACACAGCGCCCGG
1301 1306
|-----|
1311 GTGGCAGCGGTGGCTTGCGAGCTGCAGCAGCCCTGSCCAGCCTTCTCTCAATGTGAATGC
|-----| |-----|
1371 ACTGCCAATCAGACAGACAGCAGCTTGGCGACCTCGAGCAACATTAACCTGAGGAAAA
1391 1396 1404
|-----|
1431 ATTGAANAAGGGGCCCTTGGTGGGGCTTCTTATTGTGGAACCTGTTATGAAAGGAG
1442 1447
|-----|
1491 CCCCATC
...
2556 AATTCGGCCAAAGAGCGCTAGCCTCCGATTACTAAACCCCTTGGCCCAACAGTCCACA
2606 2611
|-----|
2616 TTGACGAGCCTCTTTTAGTAACCTGCTCCCGGTAATTCCTTCAGA
...
2728 CCCCTTTCAGATGCTTTAAGCTGAGAGCTTAAACCACAGGTACCATGGCTGACGCTGCC
2778 2783
|-----|
2788 AGGTTTCTGCTCCAGATAATCTATGATGGGAGGGGCATATTTTTTA
...
3749 TCCTCTCCCATTCACACCTTAGAATGGAGTCTTCTGGGGGAGCTGCAAGGTTCTCC
3799 3804
|-----|
3809 CTAGGACAGATGGAGCCTCCCTTTCTCTCATCTACTCTGTGTGGTGGT
...

4 matches found in sequence:
aaal6678 ; Human secreted protein clone LL89_3 nucleotide sequence SEQ ID NO:1
(from "mycobacterng.seq")

| | |
|----|--|
| AC | AA16676; |
| XX | |
| XX | 16-JUN-2000 (first entry) |
| XX | |
| DE | Human secreted protein clone ko258_4 nucleotide sequence SEQ ID NO:117. |
| XX | |
| XX | Human; secreted protein; immunestimulant; immunesuppressant; virucide; |
| KW | antibacterial; antifungal; cytostatic; antiinflammatory; dermatological; |
| KW | antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide; |
| KW | antithyroid; immune deficiency; severe combined immunodeficiency; SCID; |
| KW | infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus; |
| KW | connective tissue disease; multiple sclerosis; erythematosis; |
| KW | rheumatoid arthritis; autoimmune pulmonary inflammation; asthma; |
| KW | Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis; |
| KW | insulin dependent diabetes mellitus; graft-versus-host-disease; |
| KW | autoimmune inflammatory eye disease; allergy; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| XX | WO200009552-A1. |
| PN | |
| XX | |
| PD | 24-FEB-2000. |
| XX | |
| PF | 13-AUG-1999; 99WO-US18298. |
| XX | |
| PR | 14-AUG-1998; 98US-0096622. |
| PR | 17-AUG-1998; 98US-0096815. |
| PR | 04-SEP-1998; 98US-0099229. |
| PR | 23-OCT-1998; 98US-0103368. |
| PR | 08-JAN-1999; 99US-0115234. |
| PR | 12-FEB-1999; 99US-0119931. |
| PR | 18-FEB-1999; 99US-0120575. |
| PR | 30-APR-1999; 99US-0132020. |
| PR | 11-AUG-1999; 99US-0096622. |
| PA | |
| XX | (GEMY) GENETICS INST INC. |
| PI | Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C; |
| PI | Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V; |
| PI | Wong GG, Clark HF, Fichtel K; |
| PI | WPI: 2000-205979/18. |
| DR | P-PSDB; AA94956. |
| XX | |
| PT | New polynucleotides encoding secreted proteins, which may have e.g. |
| PT | nutritional, chemokine, immune stimulating or suppressing, |
| PT | hematopoiesis regulating, tissue growth, activin/inhibin |
| PT | antiinflammatory or tumor inhibition activity |
| XX | |
| PS | Claim 126; Page 577-578; 641pp: English. |
| XX | |
| CC | AA16618 to AA16697 encode the human secreted proteins given in |
| CC | AA94898 to AA94980, isolated from human adult brain, adult thyroid, |
| CC | adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney |
| CC | adult placenta, adult testis, whole embryo, adult cartilage, kidney, |
| CC | foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour, |
| CC | and adult bladder, cDNA libraries. The polynucleotides and proteins are |
| CC | predicted to have biological activities which would make them suitable |
| CC | for treating, preventing or ameliorating medical conditions in humans |
| CC | and animals. The polynucleotides can be used as markers for tissues in |
| CC | which the protein is preferentially expressed, as molecular weight |
| CC | markers on Southern gels, and as chromosome markers or tags to identify |
| CC | chromosomes or to map gene positions. The proteins can be used in the |
| CC | treatment of immune deficiencies and disorders, such as severe combined |
| CC | immunodeficiency (SCID), as well as viral, bacterial, fungal and other |
| CC | infections. These infections include human immunodeficiency virus (HIV), |
| CC | hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and |
| CC | candidateis. The proteins can be used to treat autoimmune disorders such |
| CC | as connective tissue disease, multiple sclerosis, systemic lupus |
| CC | erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation, |
| CC | Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent |
| CC | diabetes mellitus, myasthenia gravis, graft-versus-host-disease and |
| CC | autoimmune inflammatory eye disease. The proteins can also be used to |

| | |
|-------------|--|
| CC | treat allergic conditions, such asthma. AAAL6698 to AAAL6774 represent |
| CC | probes for the human secreted proteins from the present invention. |
| XX | |
| SQ | Sequence 2232 BP; 484 A; 625 C; 716 G; 406 T; 1 other; |
| AAAL6676 | Length: 2232 March 5, 2002 14:18 Type: N Check: 3246 .. |
| Found using | 'seq2-3' (pappu403.key) |
| ... | |
| 524 | GAAGGAGCGCAAGCCAGGAGGAGGCTGCGCGCAGCGGAGAGCTCCGCGCGCCGAGAG 574 579 ----- |
| 584 | GCAGGAGGCCAAGAGAGCTAGGGGAGGTGCACAGCGCTGGCACCCGCGCAGGGGCAGC 619 624 ----- |
| 644 | CACCTGGCTCCGCGGCACAGGCTCACCAGGAGGCTGGACCTGGCGCGCTGCACATTGGGC 688 693 ----- |
| 704 | TAGCCTGGTCCCACGCTCTCAGGGGGGACATGCTCTCTCT |
| ... | |
| 1086 | GGGGCTGTGTGGTGGGAGAGCTCTGCTCTGCTCCCCAAGAGGCCAGTGGCGCTGCAG 1136 1141 ----- |
| 1146 | CAAGTCGCACCTCAGGCTAGACTCCAGCCAAACTCTCAACAGGAG |
| ... | |
| 1719 | ACAGCATGTGTCTCTTTTAGGAAATGCTCTTGAGGAAGTGTTCATGTGGCGCTGGTC 1769 1774 ----- |
| 1779 | AGCAGCTAGTCCCGCTTCAGGACACTGTGTCAGAGTTACCGATGAG |
| ... | |
| ----- | |
| 13 | matches found in sequence: |
| AAAL6677 | ; Human secreted protein clone kv10_8 nucleotide sequence SEQ ID NO:1 |
| (from | "mycobactering.seq") |
| TOIG | of: aaal6677 check: 4606 from: 1 to: 4086 |
| ID | AAAL6677 standard; cDNA; 4086 BP. |
| XX | |
| AC | AAAL6677; |
| XX | |
| DT | 16-JUN-2000 (first entry) |
| XX | |
| DE | Human secreted protein clone kv10_8 nucleotide sequence SEQ ID NO:119. |
| XX | |
| KW | Human; secreted protein; immunestimulant; immunosuppressant; virucide; |
| KW | antibacterial; antifungal; cytostatic; antiinflammatory; dermatological; |
| KW | antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide; |
| KW | antithyroid; immune deficiency; severe combined immunodeficiency; SCID; |
| KW | infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus; |
| KW | connective tissue disease; multiple sclerosis; erythematosis; |
| KW | rheumatoid arthritis; autoimmune pulmonary inflammation; asthma; |
| KW | Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis; |
| KW | insulin dependent diabetes mellitus; graft-versus-host-disease; |
| KW | autoimmune inflammatory eye disease; allergy; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200009552-A1. |
| XX | |
| PD | 24-FEB-2000. |
| XX | |

```
688 CACCCCGGGCCCTCGCCCTGGCTGACCGAGCTGCCGAGCTCGGGCCAAAGGCGCAGG
    694 699
748 CG
...
1103 AACTAGGACGAGGATTCTCTGGGTCTTAATCGCAGGTCCATCACTGACTGTGACGCTACG
    |-----|
    1153 1158
1163 AAGGCCCTTGGCCCTCTCTGAGCCTCGGTTTCGCGCACTCAAGTAAT
...
8 matches found in sequence:
aaal6675 ; Human secreted protein clone khl3_4 nucleotide sequence SEQ ID NO:1
(from "mycobactering.seq")
TOIG of: aaal6675 check: 5109 from: 1_ to: 843

ID AAAl6675 standard; cDNA; 843 BP.
XX
AC AAAl6675;
XX
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone khl3_4 nucleotide sequence SEQ ID NO:115.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antilasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0103668.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Fachtel K;
XX
XX WPI; 2000-205979/18.
XX P-PSDB; AAY94955.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity -
XX
XX Claim 124; Page 577; 641pp; English.
PS
```

```
XX
CC AAAl6618 to AAAl6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC foetal placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAl6698 to AAAl6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 843 BP; 186 A; 262 C; 198 G; 197 T; 0 other;

AAAl6675 Length: 843 March 5, 2002 14:18 Type: N Check: 5109 ..
Found using 'seq2-3' (pappu403.key)

1 CCAGAAATCTGGCAGCTGACGGGGACCTAGGAGCAGACGACGCGCACAAACGCCAGCTTGG
    |-----|
    36 41
61 CAGGCGCTGCCAGGCGGGTGCTCACCCTAGCTAGCTCTTCATGTATTATCATCAAAGCCTTTC
    |-----|
    63 68
121 TGTCCACCAGCGCCCATCTTCCTTCCAGCTGTGAACTGTGCCATGGGTACGCGACGCG
    |-----|
    130 135
181 CTCTCGGGCGCTCTCTTGCNAGCAGGGGACTCGCCGCGCGCGCCACCGCTCATGACGCGC
    |-----|
    187 192
241 GCGCGCTGCACGCTTTCGCGCGCTCTGTCAGGCCCGCCAGGAGCCAGAGGGGTACCTGGAGGC
    |-----|
    257 262
301 CTGGCCCGCGCCT
...
557 GGGGATATACCTGAGTTGAGCTAGAGGATTTTATCCCTGTTGGGATGGGGGAGCTGCTCGG
    |-----|
    607 612
617 GAAGTGTGTTTCTTAACATAAAGACTGCAGGAAGTGTCAACTTTTA
...
-----
5 matches found in sequence:
aaal6676 ; Human secreted protein clone ko258_4 nucleotide sequence SEQ ID NO:
(from "mycobactering.seq")
TOIG of: aaal6676 check: 3246 from: 1 to: 2232
ID AAAl6676 standard; cDNA; 2232 BP.
XX
```


XX PS Claim 118; Page 571-572; 641pp; English.

XX CC AAA16618 to AAA16697 encode the human secreted proteins given in

CC CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,

CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,

CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,

CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,

CC and adult bladder, cDNA libraries. The polynucleotides and proteins are

CC predicted to have biological activities which would make them suitable

CC for treating, preventing or ameliorating medical conditions in humans

CC and animals. The polynucleotides can be used as markers for tissues in

CC which the protein is preferentially expressed, as molecular weight

CC markers on Southern gels, and as chromosome markers or tags to identify

CC chromosomes or to map gene positions. The proteins can be used in the

CC treatment of immune deficiencies and disorders, such as severe combined

CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such

CC as connective tissue disease, multiple sclerosis, systemic lupus

CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and

CC autoimmune inflammatory eye disease. The proteins can also be used to

CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent

CC probes for the human secreted proteins from the present invention.

XX SQ Sequence 1199 BP; 221 A; 394 C; 349 G; 235 T; 0 other;

AAA16672 Length: 1199 March 5, 2002 14:18 Type: N Check: 6601 ..

Found using 'seq2-3' (pappu403.key)

...

403 AGGACTGGGAAGTGCAGTACCAACAGACACCCCGGTGGCCCGCTTTGACGCTCAATG |-----|

453 458

463 CCCCCGACCTCTACATTCACAGCAATGGCTTTCATCACCTACGTTTTT

...

747 TACCTGTGTGCTGGGTGCTGCTAGGCATCTTTGTGTTCATGATCCGGACGCTCGCG |-----|

797 802

807 CTGAAGATCTTGCACAGCAGCAGCTGAGGGGGTCCCCGGTGGCGTG

...

4 matches found in sequence:

aaa16673 ; Human secreted protein clone fy356_14 nucleotide sequence SEQ ID NO

(from "mycobacterieng.seq")

TOIG of: aaa16673 check: 4773 from: 1 to: 2024

XX ID AAA16673 standard; cDNA; 2024 BP.

AC AAA16673;

XX DT 16-JUN-2000 (first entry)

XX DE Human secreted protein clone fy356_14 nucleotide sequence SEQ ID NO:111.

XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;

KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;

KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;

KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;

KW connective tissue disease; multiple sclerosis; erythematosus;

KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;

KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;

KW KW insulin dependent diabetes mellitus; graft-versus-host-disease;

XX KW autoimmune inflammatory eye disease; allergy; ss.

OS Homo sapiens.

XX PN WO200009552-A1.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18298.

XX PR 14-AUG-1998; 98US-0096622.

PR 17-AUG-1998; 98US-0096815.

PR 04-SEP-1998; 98US-0099229.

PR 23-OCT-1998; 98US-0105368.

PR 08-JAN-1999; 99US-0115234.

PR 12-FEB-1999; 99US-0119931.

PR 18-FEB-1999; 99US-0120575.

PR 30-APR-1999; 99US-0132020.

PR 11-AUG-1999; 99US-0096622.

XX (GEMY) GENETICS INST INC.

PA Jacobs K, McCoy JW, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;

XX WPI; 2000-205979/18.

DR P-PSDB; AAY94953.

XX New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,

PT hematopoiesis regulating, tissue growth, activin/inhibin

PT antiinflammatory or tumor inhibition activity -

XX Claim 120; Page 573; 641pp; English.

PS AAA16618 to AAA16697 encode the human secreted proteins given in

XX CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,

CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,

CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,

CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,

CC and adult bladder, cDNA libraries. The polynucleotides and proteins are

CC predicted to have biological activities which would make them suitable

CC for treating, preventing or ameliorating medical conditions in humans

CC and animals. The polynucleotides can be used as markers for tissues in

CC which the protein is preferentially expressed, as molecular weight

CC markers on Southern gels, and as chromosome markers or tags to identify

CC chromosomes or to map gene positions. The proteins can be used in the

CC treatment of immune deficiencies and disorders, such as severe combined

CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such

CC as connective tissue disease, multiple sclerosis, systemic lupus

CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and

CC autoimmune inflammatory eye disease. The proteins can also be used to

CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent

CC probes for the human secreted proteins from the present invention.

XX SQ Sequence 2024 BP; 775 A; 315 C; 385 G; 548 T; 1 other;

AAA16673 Length: 2024 March 5, 2002 14:18 Type: N Check: 4773 ..

Found using 'seq2-3' (pappu403.key)

...

227 CAAAAATTGAGAGAAGCTTTTACTCTGATGATCAAGTTTTCGACGAGCTGTTATTA |-----|

277 282

XX 16-JUN-2000 (first entry)

XX Human secreted protein clone dw78_1 nucleotide sequence SEQ ID NO:107.

DE Human; secreted protein; immunostimulant; immunosuppressant; virucide;

XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;

XX antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;

XX antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;

XX connective tissue disease; multiple sclerosis; erythematosis;

XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;

XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;

XX insulin dependent diabetes mellitus; graft-versus-host-disease;

XX autoimmune inflammatory eye disease; allergy; ss.

OS Homo sapiens.

XX WO200009552-A1.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18298.

XX 14-AUG-1998; 98US-0096622.

XX 17-AUG-1998; 98US-0096815.

XX 04-SEP-1998; 98US-0099229.

XX 23-OCT-1998; 98US-0105368.

XX 08-JAN-1999; 99US-0115234.

XX 12-FEB-1999; 99US-0119931.

XX 18-FEB-1999; 99US-0120575.

XX 30-APR-1999; 99US-0132020.

XX 11-AUG-1999; 99US-0096622.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fecthel K;

XX WPI: 2000-205979/18.

DR P-PSDB; AAY94951.

XX New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,

PT hematopoiesis regulating, tissue growth, activin/inhibin

XX antiinflammatory or tumor inhibition activity

XX Claim 116; Page 570-571; 641pp; English.

XX AAA16618 to AAA16697 encode the human secreted proteins given in

CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,

CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,

CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,

CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,

CC and adult bladder. cDNA libraries. The polynucleotides and proteins are

CC predicted to have biological activities which would make them suitable

CC for treating, preventing or ameliorating medical conditions in humans

CC and animals. The polynucleotides can be used as markers for tissues in

CC which the protein is preferentially expressed, as molecular weight

CC markers on Southern gels, and as chromosome markers or tags to identify

CC chromosomes or to map gene positions. The proteins can be used in the

CC treatment of immune deficiencies and disorders, such as severe combined

CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such

CC as connective tissue disease, multiple sclerosis, systemic lupus

CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and

CC autoimmune inflammatory eye disease. The proteins can also be used to

CC treat allergic conditions, such asthma. AAA16698 to AAA16774 represent

CC probes for the human secreted proteins from the present invention.

XX Sequence 1418 BP; 354 A; 350 C; 356 G; 358 T; 0 other;

XX AAA16671 Length: 1418 March 5, 2002 14:18 Type: N Check: 4674

XX Found using 'seq2-3' (pappu403.key)

XX

526 CACTCCCTGCCTGCCTGTTCCCTATATCCCCCAGGCTGGATCTTCAGGCTCGGTG |-----|

576 581

586 TGGGAGGGGCCCTGCCTGCTTGCCTTGATTGTTGCTCCCTGGG

XX

2 matches found in sequence:

aaa16672 ; Human secreted protein clone fh116_11 nucleotide sequence SEQ ID NO

(from "mycobacterng.seq")

TOIG of: aaa16672 check: 6601 from: 1 to: 1199

ID AAA16672 standard; cDNA; 1199 BP.

XX

AC AAA16672;

XX

DT 16-JUN-2000 (first entry)

XX

DE Human secreted protein clone fh116_11 nucleotide sequence SEQ ID NO:109.

XX

XX Human; secreted protein; immunostimulant; immunosuppressant; virucide;

XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;

XX antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;

XX antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;

XX connective tissue disease; multiple sclerosis; erythematosis;

XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;

XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;

XX insulin dependent diabetes mellitus; graft-versus-host-disease;

XX autoimmune inflammatory eye disease; allergy; ss.

XX

OS Homo sapiens.

XX WO200009552-A1.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18298.

XX 14-AUG-1998; 98US-0096622.

XX 17-AUG-1998; 98US-0096815.

XX 04-SEP-1998; 98US-0099229.

XX 23-OCT-1998; 98US-0105368.

XX 08-JAN-1999; 99US-0115234.

XX 12-FEB-1999; 99US-0119931.

XX 18-FEB-1999; 99US-0120575.

XX 30-APR-1999; 99US-0132020.

XX 11-AUG-1999; 99US-0096622.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fecthel K;

XX WPI: 2000-205979/18.

DR P-PSDB; AAY94952.

XX New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,

PT hematopoiesis regulating, tissue growth, activin/inhibin

PT antiinflammatory or tumor inhibition activity


```

...
119 AGTGGGTGGTGGATCTCGGCTCACTGAAACCTCTGCTCCTGGGTTCAAGCGATTCTC
      |-----|
      169 174

179 CTGCTTCACCTCCGAGTAGCTGGGATTACAGGTGTCACCAACCAC
      ...
-----
28 matches found in sequence:
aaal6669 : Human secreted protein clone dd428_4 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6669 check: 1731 from: 1 to: 1530

ID   AAA16669 standard; cDNA; 1530 BP.
XX
AC   AAA16669;
XX
DT   16-JUN-2000 (first entry)
XX
XX   Human secreted protein clone dd428_4 nucleotide sequence SEQ ID NO:103.
XX
XX   Human: secreted protein; immunestimulant; immunesuppressant; virucide;
XX   antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
XX   antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
XX   antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
XX   infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
XX   connective tissue disease; multiple sclerosis; erythematosis;
XX   rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
XX   Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
XX   insulin dependent diabetes mellitus; graft-versus-host-disease;
XX   autoimmune inflammatory eye disease; allergy; ss.
XX
OS   Homo sapiens.
XX
XX   WO200009552-A1.
XX
XX   24-FEB-2000.
XX
XX   13-AUG-1999; 99WO-US18298.
XX
XX   14-AUG-1998; 98US-0096622.
XX   17-AUG-1998; 98US-0096815.
XX   04-SEP-1998; 98US-0099229.
XX   23-OCT-1998; 98US-0105368.
XX   08-JAN-1999; 99US-0115234.
XX   12-FEB-1999; 99US-0119931.
XX   18-FEB-1999; 99US-0120575.
XX   30-APR-1999; 99US-0132020.
XX   11-AUG-1999; 99US-0096622.
XX
XX   (GENY ) GENETICS INST INC.
XX
XX   Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX   Merberg D, Treacy MJ, Agostino MJ, Steininger RJ, Spaulding V;
XX   Wong GG, Clark HF, Fechtel K;
XX
XX   WPI: 2000-205979/18.
XX   P-PSDB: AAY94949.
XX
XX   New polynucleotides encoding secreted proteins, which may have e.g.
XX   nutritional, chemokine, immune stimulating or suppressing,
XX   hematopoiesis regulating, tissue growth, activin/inhibin
XX   antiinflammatory or tumor inhibition activity
XX
XX   Claim 112; Page 568; 641pp; English.
XX
XX   AAA16618 to AAA16697 encode the human secreted proteins given in
XX   AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX   adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney.

```

```

CC   adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC   foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC   and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC   predicted to have biological activities which would make them suitable
CC   for treating, preventing or ameliorating medical conditions in humans
CC   and animals. The polynucleotides can be used as markers for tissues in
CC   which the protein is preferentially expressed, as molecular weight
CC   markers on Southern gels, and as chromosome markers or tags to identify
CC   chromosomes or to map gene positions. The proteins can be used in the
CC   treatment of immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC   infections. These infections include human immunodeficiency virus (HIV),
CC   hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC   candidiasis. The proteins can be used to treat autoimmune disorders such
CC   as connective tissue disease, multiple sclerosis, systemic lupus
CC   erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC   Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC   diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC   autoimmune inflammatory eye disease. The proteins can also be used to
CC   treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC   probes for the human secreted proteins from the present invention.
XX
XX   Sequence 1530 BP; 490 A; 336 C; 351 G; 353 T; 0 other;
XX
AA16669 Length: 1530 March 5, 2002 14:18 Type: N Check: 1731
Found using 'seq2-3' (pappu403.key)
...
89   AATGTGCACATGTACCTAAACTTAAAGTATATATATATATATAAAAAAGACATCGCTG
      |-----| |-----| |-----|
      139 144

149   AGCAGCGTGTATACGCATCGCTAATGAGGACACCATACAGGCATCGCTAACGATCAGC
      |-----| |-----| |-----| |-----|
      166 171 193 198 205 209 252 259 264

209   CTGTACACACATCACTAATGATGACACCGCTATAAGACATCGCTAATATGACGCTGTAT
      |-----| |-----| |-----|
      210 247 252 259 264

269   ACGACATCGCTAATGACACCGCTACAGGCACGCTAACGAGGATGCTGTACACGACATCAC
      |-----|
      274 279

329   TAATGAGGACAGTGTACAGCCATCACTAATGAGGACACTGTATATGGCATCGCTAACGA
      |-----|
      378 383

389   GGACACTGTACAGGCATTGCTTAACGAGGAGCGCTGTACACACATCGCTAATGAGGACAC
      |-----| |-----| |-----|
      417 422 432 437

449   CATATAAGACATCACCAATGAGGATGCTGTATATGACATCGCTAATGGCCACCCACAAGGC
      |-----|
      486 491

509   ATGCTAACGAGGAGCGCTGTAGACGACATTCGCTTTATAAGGACACCGCTACAGACATCGCTA
      |-----| |-----| |-----| |-----|
      520 525 529 534 562 567

569   ACAGGACGCTGTATACGACATCGCTAATGAGGACGCTGTATATGACATCGCTAATGAGG
      |-----| |-----| |-----| |-----|
      574 579 589 594 601 606 616 621

629   ATGCTTTACAAGACATAGTAGTGAATGAGGTTCGTGTATATGACATCGCTAATGAGGACATTG
      |-----|
      670 675

689   TATATGACATCGCTAATGAGGAGCGCTCTATACGACATCACTAATGAGGAGCGCTGTATACA
      |-----| |-----| |-----|

```

PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
XX Claim 108; Page 566; 64lpp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAAY94898 to AAAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 915 BP; 206 A; 290 C; 249 G; 170 T; 0 other;

AAA16667 Length: 915 March 5, 2002 14:18 Type: N Check: 3307 ..
Found using 'seq2-3' (pappu403.key)

...
196 GTCTGAGGGTGCTGCCTGTCTATGCGGGGAGCCATCTCCAGGGGGCCCTCATCGCCATCG
|-----| |---|
246 251
252

256 -|
TCTCCACAGGTCTCGTGGCTTCTTGCTGCTGCTGCTGGTATCCTCTGCTGGGCGCT
257

316 GCCATTCTGCTGCGCGAGCTGACACTCTCTCTGATCATCAGTCCCACTCCAGCCCTG
|-----|
333 338

376 GCCCCTGCTCCTGA

1 match found in sequence:
aaal6668 : Human secreted protein clone cw1475_2 nucleotide sequence SEQ ID NO
(from "mycobacteng.seq")
TOIG of: aaal6668 check: 288 from: 1 to: 2915

ID AAA16668 standard; cDNA; 2915 BP.
XX
AC AAA16668;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone cw1475_2 nucleotide sequence SEQ ID NO:101.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;

KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX Homo sapiens.
OS
XX
XX W0200009552-A1.
XX
PD 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAAY9498.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
XX Claim 110; Page 566-567; 64lpp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAAY94898 to AAAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2915 BP; 888 A; 492 C; 536 G; 999 T; 0 other;

AAA16668 Length: 2915 March 5, 2002 14:18 Type: N Check: 288 ..
Found using 'seq2-3' (pappu403.key)

insulin dependent diabetes mellitus; graft-versus-host-disease;
autoimmune inflammatory eye disease; allergy; ss.
Homo sapiens.
WO200009552-A1.
24-FEB-2000.
13-AUG-1999; 99WO-US18298.
14-AUG-1998; 98US-0096622.
17-AUG-1998; 98US-0096815.
04-SEP-1998; 98US-0099229.
23-OCT-1998; 98US-0105368.
08-JAN-1999; 99US-0115234.
12-FEB-1999; 99US-0119931.
18-FEB-1999; 99US-0120575.
30-APR-1999; 99US-0132020.
11-AUG-1999; 99US-0096622.
(GEMY) GENETICS INST INC.
Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
Wong GG, Clark HF, Fechtel K;
WPI: 2000-205979/18.
P-PSDB: AAY94946.
New polynucleotides encoding secreted proteins, which may have e.g.
nutritional, chemokine, immune stimulating or suppressing,
hematopoiesis regulating, tissue growth, activin/inhibin
antiinflammatory or tumor inhibition activity
Claim 106; Page 565; 641pp: English.
AAAL6618 to AAAL6697 encode the human secreted proteins given in
AAY94998 to AAY94980, isolated from human adult brain, adult thyroid,
adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
adult placenta, adult testis, whole embryo, adult cartilage, kidney,
foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
predicted to have biological activities which would make them suitable
for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
SQ Sequence 1299 BP; 491 A; 289 C; 262 G; 257 T; 0 other;

AAAL6666 Length: 1299 March 5, 2002 14:18 Type: N Check: 3881
Found using 'seq2-3' (pappu403.key)

694 GAGAACTACATTAATGAGACACAGAGCCCGAGAGTGAACAGAGAGAGAGTGGAG
744 749
|-----|

754 TTCACAAAGAACTTCCAGGATATGGCTATACCAAAAAAATCATCTTGGCGCTTAATTGTG |-----|
801 806
814 ACTGGAATACATGATTTGATTATATCTTCTGCTTCATTTGAGATCTGTTGTACCGGA |-----|
825 830
874 AGGTCAT
...
1034 CCAGCGGCCACACATCCACAGAGGCGGCGGAGACAGCAAGAGGCCATCGTTCTGTG |-----|
1084 1089
1094 CCCTTGTTCACAGTCTAATTAGTCCACCAGACCTGAAACATATG
...

3 matches found in sequence:
aaal6667 : Human secreted protein clone cw1292_8 nucleotide sequence SEQ ID NO
(from "mycobacterng.seq")
TOIG of: aaal6667 check: 3307 from: 1 to: 915
ID AAAL6667 standard; cDNA; 915 BP.
XX AC
XX AAAL6667;
XX DT
XX 16-JUN-2000 (first entry)
XX DE
XX Human secreted protein clone cw1292_8 nucleotide sequence SEQ ID NO:99.
XX KW
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
XX antidiabetic; antiasthmatic; antiarthritic; severe combined immunodeficiency;
XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
XX connective tissue disease; multiple sclerosis; erythematosus;
XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
XX insulin dependent diabetes mellitus; graft-versus-host-disease;
XX autoimmune inflammatory eye disease; allergy; ss.
XX OS
XX Homo sapiens.
XX WO200009552-A1.
XX PN
XX 24-FEB-2000.
XX PD
XX 13-AUG-1999; 99WO-US18298.
XX PF
XX 14-AUG-1998; 98US-0096622.
XX PR
XX 17-AUG-1998; 98US-0096815.
XX PR
XX 04-SEP-1998; 98US-0099229.
XX PR
XX 23-OCT-1998; 98US-0105368.
XX PR
XX 08-JAN-1999; 99US-0115234.
XX PR
XX 12-FEB-1999; 99US-0119931.
XX PR
XX 18-FEB-1999; 99US-0120575.
XX PR
XX 30-APR-1999; 99US-0132020.
XX PR
XX 11-AUG-1999; 99US-0096622.
XX PA
XX (GEMY) GENETICS INST INC.
XX PI
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX PI Wong GG, Clark HF, Fechtel K;
XX XX
XX WPI: 2000-205979/18.
XX DR
XX P-PSDB: AAY94947.
XX PT
XX New polynucleotides encoding secreted proteins, which may have e.g.

```
1107 CTTTGAATTTCAGACTTCTATGATTTTAGGAGTAGTATCCAGA
...
1216 TGATGATGAACCACTGAATGATTCTCGCTTCTGTGGTGCCAGAGTGGGTACGCTCCCTT
1266 1271
1276 GATGAGATACTACAACGCGATTGGCTTGTCAAGAGCTGTGGCAGTGGCCAAAATCG
1293 1298
1336 GAAGGCCGTGGCG
...
1459 AAGTTGTTTCCATTATGTGTGCCATGTGGATTCTCTCTAGGGGAAACGTACGCTCTGG
1509 1514
1519 CTCCAGGGACTTCCATTACCACTTGATATGTTTAAGGCAATAAGCC
...
1 match found in sequence:
aaal6665 : Human secreted protein clone bk343_2 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6665 check: 6068 from: 1 to: 1395

ID AAAl6665 standard; cDNA; 1395 BP.
XX
AC AAAl6665;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone bk343_2 nucleotide sequence SEQ ID NO:95.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-AL.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMV ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
```

```
XX
DR WPI; 2000-205979/18.
XX P-PSDB; AAY94945.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 104; Page 564; 641pp; English.
XX
CC AAAl6618 to AAAl6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependant
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAl6698 to AAAl6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1395 BP; 402 A; 261 C; 292 G; 440 T; 0 other;

AAAl6665 Length: 1395 March 5, 2002 14:18 Type: N Check: 6068
Found using 'seq2-3' (pappu403.key)

...
24 GAAGTGGCTGGCTCATTTAAGATGAGGCTTCTGCTCTCTCTAGTGGCGGCTGCGG
|-----|
74 79
84 ATGTGTCGGAGGAGCGCTCGGCCAATCTGGGCGGCGTCCACGCA
...
4 matches found in sequence:
aaal6666 : Human secreted protein clone cd205_2 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6666 check: 3881 from: 1 to: 1299

ID AAAl6666 standard; cDNA; 1299 BP.
XX
AC AAAl6666;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone cd205_2 nucleotide sequence SEQ ID NO:97.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
```

1555 1560

1565 TTTTCTCACTCAACTTGAGATAGACTAGTTTGTACACCATTTGCCTT

...

13 matches found in sequence:

aaal6664 : Human secreted protein clone bf157_16 nucleotide sequence SEQ ID NO
(from "mycobacterng.seq")
TOTG of: aaal6664 check: 4285 from: 1 to: 2922

ID AAAl6664 standard; cDNA; 2922 BP.

AC AAAl6664;

XX 16-JUN-2000 (first entry)

XX Human secreted protein clone bf157_16 nucleotide sequence SEQ ID NO:93.

XX Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.

XX Homo sapiens.

OS

XX WO200009552-A1.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18298.

XX 14-AUG-1998; 98US-0096622.

XX 17-AUG-1998; 98US-0096815.

XX 04-SEP-1998; 98US-0099229.

XX 23-OCT-1998; 98US-0105368.

XX 08-JAN-1999; 99US-0115234.

XX 12-FEB-1999; 99US-0119931.

XX 18-FEB-1999; 99US-0120575.

XX 30-APR-1999; 99US-0132020.

XX 11-AUG-1999; 99US-0096622.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

XX Wong GG, Clark HF, Fachtel K;

XX WPI: 2000-205979/18.

XX P-PSDB: AAY94944.

XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX Claim 102; Page 561-562; 641pp; English.

XX AAAl6618 to AAAl6697 encode the human secreted proteins given in

CC AAY94988 to AAY94980, isolated from human adult brain, adult thyroid,

CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,

CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,

CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,

CC and adult bladder, cDNA libraries. The polynucleotides and proteins are

CC predicted to have biological activities which would make them suitable

CC for treating, preventing or ameliorating medical conditions in humans

CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAl6698 to AAAl6774 represent
CC probes for the human secreted proteins from the present invention.

XX Sequence 2922 BP; 828 A; 598 C; 757 G; 739 T; 0 other;

AAAl6664 Length: 2922 March 5, 2002 14:18 Type: N Check: 4285 ..
Found using 'seq2-3' (pappu403.key)

1 GAGCGGGTTAAGGCTCTGAGGGTCTTGTGGGGCCACGGCGCTGATCACCAGGTGTTGGC
37 42
|-----|

61 TTGTCGGTCTCTTATTTCCTCGCCTGGCAATGGCGACGTACACCTGCATCACTTGCCTGGGT
91 96
|-----|

121 |-----|
GGCGTTCGCGACGCGGACATGCAGCGGGCCCACTATAAGACGGACTGGCACCGCT
121 126

276 AGGCTCGGCCACCTACTGTCACCGCTTTCAGTAAGAAGTTTGCCTCTTCAACGCTTACG
326 331
|-----|

336 AGAACCACTCAAGTCCGGCGTCACTGTTGAGCTGGAGAGAGCGCGTGACGACGAGTGA
354 359
|-----|

396 ATCGGAAAGTGGAG

476 GCGGCCATCCAGCAGGCCATCAAGGCCCGCCATGCTCTCCCAAGAGCGGCCCA
526 531 535
|-----|
540
CGCGCTCAAGAGGCGCAGGAATGTCGTGGCCCGTGGTACTGTTGGCCGTGGGA

660 CAAAGCAGCAGGAGGAGGACACCGCAGGAGGAGGAGGACCTGGATGGAGACCATGTGG
710 715
|-----|

720 AAGATATTGATTCTGTGAAGAAATTGGAATGTGAGGATCTGAGCAATGACCATGTGG
770 775
|-----|

780 TGGAGCAGGATGCAGAGGAGGAAGAGGCTGAGGAAGGCCCAACCCCT

1047 CTGTACAGGCACATATGAATGACAAAGCCACTGTAGACTCTTTCACAGATGGCGATGCTG
1097 1102
|-----|

SQ Sequence 2529 BP; 621 A; 568 C; 614 G; 726 T; 0 other;
AA16662 Length: 2529 March 5, 2002 14:18 Type: N Check: 2010
Found using 'seq2-3' (pappu403.key)
...
519 TGAGAACTTTTGTGATCAGACATTGTCAGACACCGACACCTGCAGCTTTGAGCGCTTTGT
|-----|
569 574
579 GGTAGTCCCGGGGAGAACATGCACCAAGTGGCTGATGGCTCTGTG
...
1125 AGTGCAATGGCGGTGATCCGCTCACTGCAACCTCCACCTCGCGGGTTTAAGCGATTCTT
|-----|
1175 1180
1185 CTGCTCAGCCTCCCTAGTAGTAGTGGGATTACAGGTGCCACCA
...
1604 TGACCTCAGGTGATCTGCCTGCCTCGCTCCCAAGTGTGATTACAGACGTACGCC
|-----|
1654 1659
1664 ACCATGCTGGCCTGAACCTTTTATTAGTTAAAGTTGAATCCATC
...

4 matches found in sequence:
aa16663 ; Human secreted protein clone ytl4_1 nucleotide sequence SEQ ID NO:9
(from "mycobacterieng.seq")
TOIG of: aa16663 check: 6165 from: 1 to: 2390
ID AA16663 standard; cDNA; 2390 BP.
XX
AC AA16663;
XX
DT 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone ytl4_1 nucleotide sequence SEQ ID NO:91.
DE
XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.

XX (GEMY) GENETICS INST INC.
PA
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB: AAY94943.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 100: Page 559-560; 64lpp; English.
XX
CC AA16618 to AA16697 encode the human secreted proteins given in
CC AAY94998 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AA16698 to AA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2390 BP; 701 A; 433 C; 447 G; 809 T; 0 other;
AA16663 Length: 2390 March 5, 2002 14:18 Type: N Check: 6165
Found using 'seq2-3' (pappu403.key)
...
125 CTTCGTGTGATCAGTTATGTTATCATGACTGGTGCCTTTTATTTCAGTCTCAACGTTATTA
|-----|
175 180
185 AATCAATGATATTGACATATTATGAGGGAGAGAAGTCAATGCTGGAAGGATTGGGCTA
|
244
245 ACGCTAGTAGTAGCTGGAATGGTGGGCTCTATTCTTTTGTGGCTTATGGCTGGATT
|-----|
249
...
1018 AATTAGTATCCGGTTTTTGTAGTCTCATATTGTATCTGAAAGTAAGCTTCTTACGCTTTACT
|-----|
1068 1073
1078 TTTTAAAGTCGATGTTTTTCTTTTCTTTTGTAGAAAAATGGAAGCTTAG
...
1505 TTCTGCTCTGAGTCTTTTTTCTGCTGCTTCTTCTATCATGGAATAAATGCTAACGCTGTAT
|-----|

PA (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94941.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX Claim 96; Page 557; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1059 BP; 314 A; 220 C; 205 G; 320 T; 0 other;
AAA16661 Length: 1059 March 5, 2002 14:18 Type: N Check: 4104
Found using 'seq2-3' (pappu403.key)
...
406 CTGTTATTGGATTGTTGTTCTGGGAAGCTCTATAGACCACTGCTCAGCAGTCTCTTTA
|-----|
456 461
466 TTACAGCTTCTGCAGAACACTTCTGTTAGACTGTGAGAAGATGGCTG
...
971 CTGTTGGTTTGGGAAGTGCAGCAGCACTTCTCTATGTTGGTTTAAAAAACAACGACAACA
|-----|
1021 1026
1031 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3 matches found in sequence:
aaal6662 ; Human secreted protein clone yk251_1 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6662 check: 2010 from: 1 to: 2529
ID AAA16662 standard; cDNA; 2529 BP.
XX
AC AAA16662;
XX
DT 16-JUN-2000 (first entry)

XX Human secreted protein clone yk251_1 nucleotide sequence SEQ ID NO:89.
DE
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94942.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 98; Page 558; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX

KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEM) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Arcotino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fecthel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94940.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 94; Page 555-556; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC foetal placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1212 BP; 315 A; 291 C; 288 G; 316 T; 2 other;
AAA16660 Length: 1212 March 5, 2002 14:18 Type: N Check: 2048 ..
Found using 'seq2-3' (pappu403.key)
...
46 CTTGGACAGAGCGGGTGCACAAATCAGAGGATTAGTTGGGACCTGCCTTGGCGACCCCA

106 TGGCATCCCCAGAACCGTAACATTATGTGGCCCTTGAAGTGGCCTGTGGCATCGTCATGA 96 101
...
283 ATTCCATTCTCTCCGAAAGAGCATTGGTGGCCCTTGAAGTGGCCTGTGGCATCGTCATGA 333 338
343 CCCTTGTGCTGGCGGTCTCCCAAGATGTGGCCAACTTCTTCTACTACTGTGTGCTGGTGTGG 355 360
403 CTGTGCTCTTCTTCCACCAGCTGGTGGTGATCCTCTCAACGCTAGGCCCATGCTCTGG 442 447
463 TGTTTGAATCCTGCTCACTTGGCGGCTGCTGATT
...
958 GTGTGGCGCGCGCACGCGCACACTCACGCACACACAGCAGAGAGGGGCTGATC 1008 1013
1018 TTGAACATACTCTGCACAGGCATCCTTCCCTTTATAGATTGATTC
...
2 matches found in sequence:
aaal6661 ; Human secreted protein clone yk78_1 nucleotide sequence SEQ ID NO:8
(from "mycobacterng.seq")
TOIG of: aaal6661 check: 4104 from: 1 to: 1059
ID AAA16661 standard; cDNA; 1059 BP.
XX
AC AAA16661;
XX
XX 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone yk78_1 nucleotide sequence SEQ ID NO:87.
XX
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX

CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6658 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2625 BP; 469 A; 830 C; 728 G; 596 T; 2 other;

AAAL6658 Length: 2625 March 5, 2002 14:18 Type: N Check: 8915 ..
Found using 'seq2-3' (pappu403.key)

...
452 TTAGATTTTCAGTCCCGTCATATTCAGAGTTTGTGTTCATTTCTGCAACGCTGAGC
502 507
512 GTAGTGCCTCAGGTCGCTCGGTGCGGTGCATCTGAGTGTTCCTGGCAGCACTGCCAGGT
572 CTCCCTGGGAGCTGCGGCTCTTTGACCCTGTGACGTCCTGAGGCTGGGCCAGATGCTGCT
602 607
632 TCCTCCAGGAAGATTGTTCCTCCT
...
799 CAAGAGGCCATTTCCAGCTGGTGCAGCTGAGCTGGTGGCTGTGTCTGAGCGTCTGTTG
849 854
859 AGCCAGGTCACGGCACCCAGGTGGCGGGGGGCGAGCGCCCTGACCCAGCACGGGCTCA
894 899
919 AAGTGACTCTCCTGCTCTGCTCCAGCCACACCTGCATGCTGTGGCGGCGCTGGCGGTGG
965 970
979 AGCCTCGCCTAGCTGCCAGGTCTCTGGGGCTGCTGTGGAGATGAGTAGGAGCGTCC
1032 1037
1039 CTTTCAAGGAGAGCGGGGCTTCTCTGCTGGCGCGCACCCAGACCGCGT
...
1440 GCCAGGGCCATGGCTGAGCAGCAGCGGCCCCGACTCCCCCTGCTGTGAAGACGCTGGCA
1490 1495
1500 TGCACACACAGAGTGGGTATGAGAACCAGAGGGGTGACACCACCGGCTTCTTGGCGGAG
1560 CTGCTGAACAGCAACGTGGCCAGACGCTCATGCTCTTGGACTCGCTGCTGGAGAGCGCTG
1581 1586
1620 GCGGCTCGCCAGAAGGA
...

1705 AGTGTGGAACCCAGCGCCCCAGCTCCTCACAGCCATGATTGGCGGGCTGGACGACGGGG
1755 1760
1765 ACACCCCTCACAGCCCCAGTGGCCCTGGAGGCCATGTGGGGCTTGC
...
1921 CTATCCGCCCTCTTTTGGGCACCTTAACRAAGGTCTGCCACGGAGACTGTGAGGACGCTCTTCC
1971 1976
1981 TGGACAGGTGGTGGCGGGCTGGCGCCCTGCTGCTGCACCTGCAGGACCTCAGGCCA
2003 2008
2041 CCGTGGCAGCGCTGCAGGTTTGGCCCTGCGCATGTGTGGCCCAATCTGGCATGTGAGG
2049 2054
2101 AGCT
...
2209 TGGGCCGCTCTCTGACCCACCTGCCCTGTTCTACTTCAAGAGCAGCTGGGAGACGCTCCGAG
2259 2264
2269 CTGCTGCACCCCTGTTCCACCGGTAAAGCACCACCCCTGCCCCACCC
...
2487 TAAAGTGAATTTCCCTGGATTCAGGACTTTTTCCTGTCACCTGGTGACTCATCGTCTCTA
2537 2542
2547 GTAATTCACGGANANTTCTTAACCTGTTCCAAAGAGAGCTTAAAAAAC
...

13 matches found in sequence:
aaal6659 : Human secreted protein clone ye90_1 nucleotide sequence SEQ ID NO:8
(from "mycobacterieng.seq")
TOIG of: aaal6659 check: 211 from: 1 to: 1476
ID AAAL6659 standard; cDNA; 1476 BP.
XX
AC AAAL6659;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone ye90_1 nucleotide sequence SEQ ID NO:83.
XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX

PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94937.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 88; Page 548-549; 64lpp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2232 BP; 701 A; 388 C; 428 G; 714 T; 1 other;
AAA16657 Length: 2232 March 5, 2002 14:18 Type: N Check: 3689 ..
Found using 'seq2-3' (pappu403.key)
1 GTTCTTCATGCCCTCTAGAGACGGTTCTACAGCTCTCTTCCACCTGCATCCTCTTGGA
22 27
61 AGTCTCAGCCCATGTCG
...
171 AGCTTTTGTCATCAATTTTCAGAAATATTGGTCCAGTGCTCCACCGCAGCTACTC
221 226
231 AAAATTTTCTTTAATTTAATTTGAAGTGAAGTACAAATCTGAGTC
...
948 CTTCTCTCTTTGCCAGGAATAGACATCAATGTTTAAAGACAATGCTGGCTGGAGCCCTTTG
998 1003
|-----|
|-----|
|-----|
|-----|

1008 CATGAAGCCTGTAACTATGCAACACACAGTGGGTGTCAGGAAATTTTGCACGTTGTCCA
1057 1062
1068 GAGGTAGATCTGCTCACTCAAGTGGACGGGTGACTCCTTTGCAT
...

14 matches found in sequence:
aaal6658 : Human secreted protein clone ye78_1 nucleotide sequence SEQ ID NO:8
(from "mycobactering seq")
TOIG of: aaal6658 check: 8915 from: 1 to: 2625
ID AAA16658 standard; cDNA; 2625 BP.
XX
AC AAA16658;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone ye78_1 nucleotide sequence SEQ ID NO:81.
XX
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN W0200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94938.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 90; Page 551-552; 64lpp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable

ID AAA16656 standard; cDNA; 2848 BP.
 AC AAA16656;
 DT 16-JUN-2000 (first entry)
 XX Human secreted protein clone yell_1 nucleotide sequence SEQ ID NO:77.
 DE
 XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200009552-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 13-AUG-1999; 99WO-US18298.
 XX
 PR 14-AUG-1998; 98US-0096622.
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX
 (GENY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Werberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX
 DR WPI; 2000-205979/18.
 DR P-PSDB; AAY94936.
 XX
 PT New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity .
 XX
 PS Claim 86; Page 546; 641pp; English.
 XX
 CC AAA16618 to AAA16697 encode the human secreted proteins given in
 CC AAY9498 to AAY94980. Isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX
 SQ Sequence 2848 BP; 890 A; 497 C; 583 G; 877 T; 1 other;
 AA16656 Length: 2848 March 5, 2002 14:18 Type: N Check: 6687
 Found using 'seq2-3' (pappu403.key)
 ...
 12 CCTTGGTGTGTCATTCGTTGAAATTCGACACATCGCGGAAAGAAACGTTGTGT
 62 67
 72 GACGTGATCCTCATGTGTCAGGAAAGAAAGATACCTGCTCATCGTG
 ...
 551 GGATCAGGTTTATGATGCTGCAGTCAGGTGGTTGAATACGATGAACCTAATCGCCAGCC
 601 606
 611 ATTTATGTTGATATCCTTGTCTAAAGTCAGGTTTCTCTTATATCA
 ...
 1145 CTATGATACGAGAACTGAAAGCTGGCACACAAAGCCAGCATGTCACCGCGCTGCAG
 1195 1200
 1205 CCATGGGATGGTGAAGCAATGTCCTAATCTATGTTTGTGTGGA
 ...

 4 matches found in sequence:
 aaal6657 ; Human secreted protein clone ye72_1 nucleotide sequence SEQ ID NO:7
 (from "mycobactering.seq")
 TOIG of: aaal6657 check: 3689 from: 1 to: 2232
 ID AAA16657 standard; cDNA; 2232 BP.
 XX
 AC AAA16657;
 XX
 DT 16-JUN-2000 (first entry)
 XX
 DE Human secreted protein clone ye72_1 nucleotide sequence SEQ ID NO:79.
 XX
 KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.
 OS Homo sapiens.
 XX
 PN WO200009552-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 13-AUG-1999; 99WO-US18298.
 XX
 PR 14-AUG-1998; 98US-0096622.
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.

```

AAA16654 Length: 760 March 5, 2002 14:18 Type: N Check: 6014
Found using 'seq2-3', (pappu403.key)

...
227 GCGCGGATTGCTTTTTCATGTTTGCACACTTCCCTCCATGGCCCATCTATCCCGGCA
      277 282
      |-----|
287 TCCTCTCAAGTGGTTACCTGTTCAGAATCAAGCACAGACGACGAAGAAACCGGGAAAG
      324 329
      |-----|
347 AAAAATTAAGAGGCATGCTAAATAATTTGAGG
      ...
-----
4 matches found in sequence:
aaa16655 ; Human secreted protein clone yd218_1 nucleotide sequence SEQ ID NO:
      (from "mycobacterng.seq")
      TOIG of: aaa16655 check: 7780 from: 1 to: 875

ID AAA16655 standard; cDNA; 875 BP.
XX
AC AAA16655;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone yd218_1 nucleotide sequence SEQ ID NO:75.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN W0200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
(GEMY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtcl K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94935.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity

```

```

XX Claim 84; Page 545; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 875 BP; 161 A; 260 C; 251 G; 203 T; 0 other;

AAA16655 Length: 875 March 5, 2002 14:18 Type: N Check: 7780
Found using 'seq2-3', (pappu403.key)

...
15 TCTTGTCAAGCCTCTGTTCTCTCCTGAGTCCCTTGGGGAGCTGGCAGATGCCAGCCACCATC
      65 70
      |-----|
75 AGACAACGTGGAGGCCCTCATGGGCAATGGCTGAGGGGGCGGGCT
      ...
248 GGGTCCAGGAGTGGAGGGGGCTGCCCTGCAGGTCTTGGCCGATGGCCGCTTAC
      298 303
      |-----|
308 CCTGCTGCCCTGGGCTTTTGGCCTGAAGCAAAATTCCTGAGTGGGGG
      ...
523 CGGCCCCCAGGCCCTCAGAGTTCTGGATGCTTCCGTCGGGCTCCACAGGCATCGTCTCC
      573 578
      |-----|
583 CTTCCGCAAGTGGAGGGGGCGGCTTCCCGCAGGCATCTGAGCTCTGT
      ...
762 TTTTAAGATGAAGCAAGATGTCTGTAGGGTAATTGGCTCACATTAACTGTCCCGCACT
      812 817
      |-----|
822 GCAGGCGAGTGAAGTGTCTGAATGTAAAAAAGAAAAAAGAAAAAAGAAAAA
      ...
-----
3 matches found in sequence:
aaa16656 ; Human secreted protein clone yell_1 nucleotide sequence SEQ ID NO:7
      (from "mycobacterng.seq")
      TOIG of: aaa16656 check: 6687 from: 1 to: 2848

```

```

PT antiinflammatory or tumor inhibition activity -
PS Claim 78; Page 541-542; 641pp; English.
XX
CC AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, myasthenia gravis, graft-versus-host-disease and
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1999 BP; 580 A; 373 C; 372 G; 674 T; 0 other;

AAAL6652 Length: 1999 March 5, 2002 14:18 Type: N Check: 1447 ..
Found using 'seq2-3' (pappu403.key)
...
449 AGTGCAGTGGCAGCATCTCGGCTCACCAACCTCTGCCTCCAGGATTCAGCGATTGTC
|-----|
499 504
509 CTGCCTCAGCTACTCTCGGAGGCTGAGCGAGGAGAGATGGCGTGA
...
560 GGAGGTGGAGCCTGTAGTGACCGAGATGGCACCACCTGCCTCCAGCTCCAGCGACAG
|-----|
610 615
620 CAAGACTCCATATCAAAAAAAAAAAAAAAAAAAAAAAAAAAGATAATCCAAA
...
1127 AGACCAGAGAATTTAGAGGGAGATTGTGGAAGTCTAGGTGGTCAGATCTGTTCCC
|-----|
1177 1182
1187 TTATCACTGTAAATATTTCTGGGGGAAAAATGCTTCTGAGTTGTT
...
-----
2 matches found in sequence:
aaal6654 : Human secreted protein clone yd137_1 nucleotide sequence SEQ ID NO:
(from "mycobacter.ng.seq")
TOIG of: aaal6654 check: 6014 from: 1 to: 760

ID AAAL6654 standard; cDNA; 760 BP.
XX
AC AAAL6654;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone yd137_1 nucleotide sequence SEQ ID NO:73.

```

```

XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN W0200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEM ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fectel K;
XX
DR WPI: 2000-2059979/18.
XX P-PSDB: AAY94934.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
PS Claim 82; Page 544; 641pp; English.
XX
CC AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, myasthenia gravis, graft-versus-host-disease and
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 760 BP; 229 A; 158 C; 148 G; 225 T; 0 other;

```

```
XX PF 13-AUG-1999; 99WO-US18298.
XX PF
XX PR 14-AUG-1998; 98US-0096622.
XX PR 17-AUG-1998; 98US-0096815.
XX PR 04-SEP-1998; 98US-0099229.
XX PR 23-OCT-1998; 98US-0105368.
XX PR 08-JAN-1999; 99US-0115234.
XX PR 12-FEB-1999; 99US-0119931.
XX PR 18-FEB-1999; 99US-0120575.
XX PR 30-APR-1999; 99US-0132020.
XX PR 11-AUG-1999; 99US-0096622.
XX PA (GEM ) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fachtel K;
XX WPI: 2000-205979/18.
XX P-PSDB; AAY94931.
XX PT New polynucleotides encoding secreted proteins, which may have e.g.
XX PT nutritional, chemokine, immune stimulating or suppressing,
XX PT hematopoiesis regulating, tissue growth, activin/inhibin
XX PT antiinflammatory or tumor inhibition activity
XX PS Claim 76; Page 539-540; 64lpp; English.
XX CC AAA16618 to AAA16697 encode the human secreted proteins given in
XX CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX CC predicted to have biological activities which would make them suitable
XX CC for treating, preventing or ameliorating medical conditions in humans
XX CC and animals. The polynucleotides can be used as markers for tissues in
XX CC which the protein is preferentially expressed, as molecular weight
XX CC markers on Southern gels, and as chromosome markers or tags to identify
XX CC chromosomes or to map gene positions. The proteins can be used in the
XX CC treatment of immune deficiencies and disorders, such as severe combined
XX CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX CC infections. These infections include human immunodeficiency virus (HIV),
XX CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX CC candidiasis. The proteins can be used to treat autoimmune disorders such
XX CC as connective tissue disease, multiple sclerosis, systemic lupus
XX CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
XX CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
XX CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
XX CC autoimmune inflammatory eye disease. The proteins can also be used to
XX CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
XX CC probes for the human secreted proteins from the present invention.
XX SQ Sequence 2336 BP; 620 A; 508 C; 465 G; 742 T; 1 other;
XX
XX AAA16651 Length: 2336 March 5, 2002 14:18 Type: N Check: 4332
XX Found using 'seq2-3' (pappu403.key)
XX
XX 37 CTATGTCATAGAAAAACCTGGAAGGCGCTACCAATAATATAGATGATCGTCTCTG
XX 87
XX 92
XX 97 CTTGCTGAGCTAACAGGGGTGTCAAGCTTCCATTGTTGGTATCTACT
XX
XX
XX 1324 GGTTCGTTTCATCTTTATTTGAACTGGCTAGAAAAACAAAGTTAATTAATTGTCGTCATAC
XX 1374 1379
XX
XX 1384 CGTGGTTACCTGCATACCTCTTGGTTGTTATCAACTGTTGTTGGTGATC
XX
XX
XX 1746 AAATATCTTTTCTGGTTTACTCTTTTTTTTGGAGATGGAGTCTCACTCTGTCGCCCAGG
XX 1796 1801
XX
XX 1806 CTGGAGTGCAGTGCCAGATCTCAGCTCACTGTAAGCTCCACCTCCCGGGTTTCATGCCAT
XX
XX 1866 TCTTCTGCCTCAGAGAGTAGCTGGGACTACAGCGCGCCGCCACCATGCCTGGCTAATTTT
XX 1897 1902
XX
XX 1926 TTTTCTTTTCTTTTTTCTTTTTTGT
XX
XX
XX 3 matches found in sequence:
XX aaal6652 ; Human secreted protein clone rd432_4 nucleotide sequence SEQ ID NO:
XX (from "mycobacterng.seq")
XX TOIG of: aaal6652 check: 1447 from: 1 to: 1999
XX
XX ID AAA16652 standard; cDNA; 1999 BP.
XX AC AAA16652;
XX XX
XX DT 16-JUN-2000 (first entry)
XX XX
XX DE Human secreted protein clone rd432_4 nucleotide sequence SEQ ID NO:69.
XX XX
XX KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
XX KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
XX KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
XX KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
XX KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
XX KW connective tissue disease; multiple sclerosis; erythematosus;
XX KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
XX KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
XX KW insulin dependent diabetes mellitus; graft-versus-host-disease;
XX KW autoimmune inflammatory eye disease; allergy; ss.
XX OS Homo sapiens.
XX XX
XX PN W0200009552-A1..
XX XX
XX PD 24-FEB-2000.
XX XX
XX PF 13-AUG-1999; 99WO-US18298.
XX XX
XX PR 14-AUG-1998; 98US-0096622.
XX PR 17-AUG-1998; 98US-0096815.
XX PR 04-SEP-1998; 98US-0099229.
XX PR 23-OCT-1998; 98US-0105368.
XX PR 08-JAN-1999; 99US-0115234.
XX PR 12-FEB-1999; 99US-0119931.
XX PR 18-FEB-1999; 99US-0120575.
XX PR 30-APR-1999; 99US-0132020.
XX PR 11-AUG-1999; 99US-0096622.
XX XX
XX PA (GEM ) GENETICS INST INC.
XX XX
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX PI Wong GG, Clark HF, Fachtel K;
XX XX
XX WPI: 2000-205979/18.
XX DR
XX DR P-PSDB; AAY94932.
XX XX
XX PT New polynucleotides encoding secreted proteins, which may have e.g.
XX PT nutritional, chemokine, immune stimulating or suppressing,
XX PT hematopoiesis regulating, tissue growth, activin/inhibin
```

PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEM) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94930.
XX
DR New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 74; Page 537; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1558 BP; 282 A; 492 C; 411 G; 373 T; 0 other;

AAA16650 Length: 1558 March 5, 2002 14:18 Type: N Check: 7828 ..
Found using 'seq2-3' (pappu403.key)

```
1      GTGCACAGGGCCGACGTAGATCTCGCGCCACACAGGGGGCCGAGCATGTGTCGGCGGCGG
      25 30
      |-----|
61      CGGGCGCTCCGGGAGGCCAGGACAGCTGATGGTTGTGGCAGAAACATCTCAAGGTAG
      64 69
      |-----|
      ...
181     TGGCTCCCTCCCTCATGACCGCTGGATCCTCCTGCTGTCAGCTGTCAAGGTTCTCC
      231 236
      |-----|
241     ATCACTGGCATATGGACTGTGTATGCCATGGCTGTGATGAACCAACC
      ...
326     GCCCTCCTGACCCCTGCTGAGCAAGGGGGTCCCAAGACCTGCTGCACCCCTGGACAGTGCC
      376 381
      |-----|
```

```
386      CCCTCATCAGCAAGTGTGGCTCTCTATCCCCGAGAAGCTGCCTCTT
      ...
506      AGAGTGGCACTCTTGGGTTTAACACCACGCGCACTCATCACAGGCTGCACCAACGCTCGG
      556 561
      |-----|
566      GCCTCTTGGTGGTTGGCAACTTTTCAGGTGGATCATGCCAGGTCTCT
      ...
694      CCGCCCCGCTGGACCTGGCTGTGGCTATCTCGCAAGTGTGCTGGCTGTCTATCGCCTTTA
      744 749
      |-----|
754      TCACCTGGTCTCTCAGTGGAGTCTCTTTGTTCATGAGAGTTCTCAGCTGCAACATGGG
      ...
814      CAGCCCTGTGTGAGTGGGTGTGTCTATCCATATCGTCATTTTCTATGGCACCTTCAGT
      846 851
      |-----|
874      ACGAGTTTGGGCGAGTCTCTCTCAGACAC
      ...
945      TCCTCCGGGAGCAGCAGCACTCCACCACCTCAACTGTGCCCGCGAGAGCATCGCTATGA
      995 1000
      |-----|
1005     TCTAAGGTCTGGGAGGGTGGCTGGCCGGCTCCACAGCACCCAC
      ...
1173     TGGCGCCTGCCAGCTGCCACAGCTGCATGACCTTCCACCCACCGGTGCTGTTTGT
      1223 1228
      |-----|
1233     TTTTAAAGGTCACCTGTCTCTCACTCACCAGCCAGCCCTTCAGGTG
      ...
-----
4 matches found in sequence:
aaal6651 ; Human secreted protein clone qyl261_2 nucleotide sequence SEQ ID NO
(from "mycobacterieng.seq")
TOIG of: aaal6651 check: 4332 from: 1 to: 2336

ID      AAA16651 standard; cDNA; 2336 BP.
XX      XX
AC      AAA16651;
XX      XX
DT      16-JUN-2000 (first entry)
XX      XX
DE      DE
XX      XX
KW      Human secreted protein clone qyl261_2 nucleotide sequence SEQ ID NO:67.
KW      Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW      antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW      antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW      antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW      infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW      connective tissue disease; multiple sclerosis; erythematosis;
KW      rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW      Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW      insulin dependent diabetes mellitus; graft-versus-host-disease;
KW      autoimmune inflammatory eye disease; allergy; ss.
XX      XX
OS      Homo sapiens.
XX      XX
PN      WO200009552-A1.
XX      XX
PD      24-FEB-2000.
```

```

...
-----
2 matches found in sequence:
aaal6649 ; Human secreted protein clone pw460_5 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6649 check: 3164 from: 1 to: 1812

ID AAAL6649 standard; cDNA; 1812 BP.
XX
AC AAAL6649;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone pw460_5 nucleotide sequence SEQ ID NO:63.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JW, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94929.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 72; Page 536; 641pp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the

```

```

CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependant
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1812 BP; 586 A; 293 C; 343 G; 590 T; 0 other;

AAAL6649 Length: 1812 March 5, 2002 14:18 Type: N Check: 3164
Found using 'seq2-3' (pappu403.key)

...

612 AAATGTGTGCATATTCTGTTCTTGTAAATAAGATTCCAGAGCTCTTTGATCGCTTTTA
662 667
|-----|

672 TAACTGCAGTTCATTTTAAATGAAGGCCAGCATATATATCTTGCAAG
|-----|
891 896

841 AGCAGGGCATAGGAGAAATGTCAAGTAGTCTAATGCAATTTTCACAGACGCTTCGG
891 896
|-----|

901 GAAATATTTCATGCTGCCCATCTGTTTCATTTCTAAATTTATATTCA
|-----|

...
-----
9 matches found in sequence:
aaal6650 ; Human secreted protein clone qal36_1 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6650 check: 7828 from: 1 to: 1558

ID AAAL6650 standard; cDNA; 1558 BP.
XX
AC AAAL6650;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone qal36_1 nucleotide sequence SEQ ID NO:65.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JW, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94929.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 72; Page 536; 641pp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the

```

```

2607 CCTTGACGGATTCTCCAGGCGGACAGAGGGCTGGGAAGATGGCTGGTGGAGAGCTT
      |-----|
      2637 2642

2667 AACAGAGGAACCTCAGAAGATTCTG

...

2703 CCCCACCCCCACCAGCGGCACAGATTGTACTACCGGGGAGAGGCATCCCTGGCGCTGTCT
      |-----|
      2753 2758

2763 CCCACTGGACAGAGAGGCTGGCCATGGGCGCCAGGGGTACAGGCC

...

2846 GTTTTCTCATTTGTTGGTGTGTGGGTGGGCGAGGGGTAGGGCGGAGAGGATGCTT
      |-----|
      2896 2901

2906 GGATTTTGTGTTCTATTAGAACACACAGTTTGTCTTAATTCA

...

-----
4 matches found in sequence:
aaal6648 ; Human secreted protein clone pg195_1 nucleotide sequence SEQ ID NO:
(from "mycobacterieng.seq")
TOIG of: aaal6648 check: 8996 from: 1 to: 3145

ID AAAL6648 standard; cDNA; 3145 BP.
XX
AC AAAL6648;
XX
DT 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone pg195_1 nucleotide sequence SEQ ID NO:61.
DE
XX
XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PE 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
(GEMY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steining RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
```

```

XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94928.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 70; Page 533-534; 64lpp; English.
XX
CC AAAL6618 to AAY16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult cartilage, kidney,
CC adult placenta, adult testis, whole embryo, adult neural, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 3145 BP; 1078 A; 599 C; 634 G; 834 T; 0 other;

AAAL6648 Length: 3145 March 5, 2002 14:18 Type: N Check: 8996
Found using 'seq2-3' (pappu403.key)

...

142 TAAAGGGGAGGTTTCATTAGGACCTGCTCACTACAGTGTGCTGCGATCAAGCTTCTAT
      |-----|
      192 197

202 TGTCAAACCTCTGTGTTATCCTTTATGTCAGAGAAGCTGGGAATT

...

687 CCAACACACAGCACAAAGAACAGACAGTAGTAGATGTCACTGAGCAGTTAGCGCATTTGCA
      |-----|
      737 742

747 AATTAGATAGTCAGAGAAAGATGCTACATGTGAACCTTCCTTTACA

...

794 AAAGTAATACTACAGAGTTCTTCAAATAGCACCTTTGCGCTGAAAGATTAAAGCGTCAGAA
      |-----|
      844 849

854 AATTCTGAAAGTGAATACAGTAGGTGAGAAATAACTCTGGTAGGCA

...

2983 CTATAAACATTCAGAGGGCTGGGCATGGTGGCTCATGCCTTTAGTCTAGCGCTTGG
      |-----|
      3033 3038

3043 AGCAGAGCGCAGGAGGACTGCTTTGAGCCCCCAGGAGTTTCAAGACCAGC
```



```
183 GATCCAGGAGAAAGAGAAAGTCCGAGAACTACATCAAGAGATAGACTGGACGACATTA |-----|
    233 238
243 TAGTATTAAACAAAGATATACAAGAGGAGATACATTAAATGCAAT
...
-----
10 matches found in sequence:
aaal6647 : Human secreted protein clone ck213_12 nucleotide sequence SEQ ID NO
(from "mycobacterng.seq")
TOIG of: aaal6647 check: 1544 from: 1 to: 3191

ID AAAL6647 standard; cDNA; 3191 BP.
AC AAAL6647;
XX
DT
XX
XX 16-JUN-2000 (first entry)
DE
XX
XX Human secreted protein clone ck213_12 nucleotide sequence SEQ ID NO:59.
KW Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Fectel K;
XX WPI: 2000-205979/18.
XX P-PSDB; AAY94927.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity
XX
XX Claim 68; Page 530-531; 64lpp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
XX AAY94988 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
```

```
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 3191 BP; 998 A; 716 C; 835 G; 642 T; 0 other;
SQ
AAAL6647 Length: 3191 March 5, 2002 14:18 Type: N Check: 1544
Found using 'seq2-3' (pappu403.key)
1 CAGAGGGCTTTTATCCATGGGGCCAAATATAACCGAGGAGGCTATGGAACACTATCGCTCAAA |-----|
51 56
61 TTGGCAGATTAACGGCAAGCATACACTCTCGTCGAGGCGGTCA
...
125 AGAAGTCCCTTCACCAAGTCCAGGAGGCCATTCTAGAAACTGTGTAAGTCGCTTTCT |-----|
175 180
185 GACGGTCAAGGGCTCTCTCATCTCCCGTTCTTCTTCCAAACCATAGCCGAGTTGAATCT
195 200
245 TCTAAG
...
1636 TCACCATGTTAAAGAGCATCACTTTGGGTCTCAGGAATGACATTACATGAACGCTTTAC |-----|
1686 1691
1696 TAAATACCTAAAGAGAGGAAGTCTTAACGAGGTAAATCGAGAGAATCA
...
1846 GGCTGAGGGAATAACAAAGATGATCTCTGTTGATCTCCGCTTGATATTGAACGCTCTTAA |-----|
1896 1901
1906 AAAACATAGGAGAGAGATCTTAAACGAGGTAAATCGAGAGAATCA
...
2253 GAGGCTGGGGCAGAGGCAACTACTCTGGGAACAATAACAACACAGCAACAGATTTTC |-----|
2303 2308
2313 AAAAAAGAAACCGGAAGAGAGTGGGACCCAGAGTACACACCAA
...
2487 GCCCAAGTGGGCCCATGACAGTTTCAGTGGGAGGAGGGAGATTGAAGACGACGAGA |-----|
2537 2542
2547 GTGGGACAGAGAACCCGAGAGAGAGGACAAATATACAGCCCAACCGAGTAGGGCCAC
```

CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 1490 BP; 295 A; 407 C; 391 G; 397 T; 0 other;
SQ
AAA16645 Length: 1490 March 5, 2002 14:18 Type: N Check: 4463 ..
Found using 'seq2-3' (pappu403.key)
...
135 TTTTCTGTGACTGTGCTGCGGTAGGCTGCTAAACATCTGGCTGAACCAAGCGTTCATC
|-----|
185 190
195 CTGACCTGAAGCAGAACCTCAGAAACCAAGTAAGGCGCTGATCAT
...
499 ACCGTTTGTGTAACTCTTTCACGCAAGTAACAATCCGCCATTACAAGGTCGTCTCC
|-----|
549 554
559 TTGATCCAGTTAACGAGTCAGAACTCTTCTCCAATCAGCAGAGAA
...
1262 TTTAAATAGATGACCCCTTCAGATCATCTGTGCTACCTCTGCCCATCAGGCGTCTACA
|-----|
1312 1317
1322 CTGTCACTCAGACACCTGTGGCATGTGGAGGAGACTGCCCTGTCTCT
...

1 match found in sequence:
aaal6646 ; Human secreted protein clone rd232_5 nucleotide sequence SEQ ID NO:
(from "mycobactering.seq")
TOIG of: aaal6646 check: 8248 from: 1 to: 4184
ID AAA16646 standard; cDNA; 4184 BP.
XX
AC AAA16646;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone rd232_5 nucleotide sequence SEQ ID NO:57.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
PN
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120375.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94926.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 66; Page 528-529; 641pp; English.
PS
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 4184 BP; 1420 A; 593 C; 745 G; 1426 T; 0 other;
SQ
AAA16646 Length: 4184 March 5, 2002 14:18 Type: N Check: 8248 ..
Found using 'seq2-3' (pappu403.key)
...
...

PN W0200009552-A1.
XX 24-FEB-2000.
PD
XX 13-AUG-1999; 99WO-US18298.
PF
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEM) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB: AAY94924.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 62; Page 525; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania sp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1207 BP; 332 A; 264 C; 334 G; 277 T; 0 other;

AAA16644 Length: 1207 March 5, 2002 14:18 Type: N Check: 7148
Found using 'seq2-3' (pappu403.key)

1 ATGGCGTCCGCGCGCGTCCGGAGCATGCGCGGACCCCGAGAGCTGTTTATGACGAG
3 8 17 22
|-----|
3
ACAGAAGCCGG
61
...
369 TGCTGTGCATTGGCTCTGTAATGCTAACAGAAGTCTGAACACCTGCCAAGCGCTGTA
419 424
|-----|

429 CTGCTTTTGTGCTTCTCTCTTTTCTGTCTGTCGTCGGGAGATCCCGA
...
729 AATGGTGAAGAGAGTGGGCGATGGTGGTCTGGAGAGAGAGCGGACAGCCCGCAGGG
|-----|
779 784
789 CAGGGAAGTCAGACCTGACACCCAGTACACGGCGCCGCAAGCGCAAG
...

3 matches found in sequence:
aaal6645 ; Human secreted protein clone rc58_1 nucleotide sequence SEQ ID NO:55
(from "mycobacterng.seq")
TOIG of: aaal6645 check: 4463 from: 1 to: 1490
ID AAA16645 standard; cDNA; 1490 BP.
XX
AC AAA16645;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone rc58_1 nucleotide sequence SEQ ID NO:55.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX W0200009552-A1.
PN
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEM) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB: AAY94925.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 64; Page 526-527; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in

CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX

SQ Sequence 5160 BP; 1594 A; 1056 C; 1051 G; 1459 T; 0 other;

AAAL6643 Length: 5160 March 5, 2002 14:18 Type: N Check: 1366 ..
Found using 'seq2-3' (pappu403.key)

```
1      GATATCTTAAGCCGGGTACGTGCGACCCACGGCTCGGGAATCGCTCAGGAAAGACACACT
      |-----|
      40 45

61      GCAGACTCCACGGGACCCTGCANTAGATGGATTG
      |-----|
      152 157

102     CACAAGGGAGAAAACGGGAGGTGACACTCTCTCGCTGGAAAGAGAGGACGACGACAA
      |-----|
      152 157

162     CAAACGAAAGGACTGGACTCCATGCCGGAAGGTATCTGGAAGTCGTG
      |-----|
      268 273

218     TATAAAACAAAAGTTTGGCGAGCTGTTAAATTGCTGCTGTGTTATTAAAGAGCGTTTCA
      |-----|
      268 273

278     AGTTTCAAGTACCAAAATGTAGCTTTTACGTTGCCAAAGGAAGTTGAG
      |-----|
      1204 1209

1154     CCCCAGTTGGCACTTTGCTCTTAGATCTGAATGCCACGGATCCAGATGAGGCGCTAATG
      |-----|
      1204 1209

1214     GGAAATTCGTATATTCCTTCACGAGTCATGTGTCCTCCCAAAATAT
      |-----|
      1684 1689

1634     CTAATCCACACTGGATAGAGAAAAGAGATCTGAGTATAGTTTGACTGTAAATCGCTGAGG
      |-----|
      1684 1689

1694     ACAGGGGGACACCCAGTCTCTCTACAGTGAACATTTTACAGTTCA
      |-----|
      2092 2097

2042     AACTGGTAAGCAATACCAACAGTTGTGCTCACCATCATTTGACGAAATGACAACGTTCTG
```

```
2102     TGCTTATAGGGCCTGCATTGCGTAATAATACGGCAGAAATCACCACAT
      |-----|
      2299 2304

2249     ATGAGGAGAATATCTTTCATATAATTGATCCAGCATGTGACATCCATCAGATGACGCTTACGA
      |-----|
      2299 2304

2309     TGGATTCTGTTCCTACACAGAAATGGGAGCTGTCAATTATCATCA
      |-----|
      2706 2711

2656     ACATTGGTGCTACCATATAAATGGCACTCTGCCCATCAGATCTCATCATCAGATGCTCTCTCCA
      |-----|
      2706 2711

2716     TCTTCATCTCTACCTTAGAAGAGGGCAGATGGGCGGCGGACAGA
      |-----|
      3082 3087

3032     ATGATTTGGGGCGAGATTTCTCCAATAGATAGGCTGTTGGGTGAAGGATTGACGCACTGT
      |-----|
      3082 3087

3092     TTCTCACAGATGGAAGAATTTCCAGCAGCTATGAGACTCTGCACGGA
      |-----|
      3286 3291

3236     AATTCCCAACGCAACCCAGCAGCATCCACATCAGAGTCTTTGAGGATGACGCTCAGC
      |-----|
      3286 3291

3296     CTGCAGATTCGGTGAAGAAGAGAGATGTTTTTCCACCTTTTGGAAAGGACTCCCAACG
      |-----|
      3357 3362

3356     -|
      ATGAGGACACTGGGATACCGACGACATCATCTCTGCTCTCGGAATGAGCAGTGTGTTC
      3357

3416     AGCGTCTCTTACCGCCTTCCCTGGACACCTATTCTGAATGACGTGAGGTGGATCGG
      |-----|
      3416 3421

      TOIG of: aaal6644 check: 7148 from: 1 to: 1207

      ID AAAL6644 standard; cDNA; 1207 BP.
      XX
      AC AAAL6644;
      XX
      DT 16-JUN-2000 (first entry)
      XX
      DE Human secreted protein clone qy338_9 nucleotide sequence SEQ ID NO:53.
      XX
      KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
      KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
      KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
      KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
      KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
      KW connective tissue disease; multiple sclerosis; erythematosis;
      KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
      KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
      KW insulin dependent diabetes mellitus; graft-versus-host-disease;
      KW autoimmune inflammatory eye disease; allergy; ss.
      XX
      OS Homo sapiens.
      XX
```

XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fachtel K;
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94922.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 58; Page 518-519; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1594 BP; 455 A; 329 C; 365 G; 445 T; 0 other;

AAA16642 Length: 1594 March 5, 2002 14:18 Type: N Check: 2777
Found using 'seq2-3' (pappu403.key)
...
52 AGACTGTGAGAAATGAGTTTTTCACTGACTTACCTCCGCGGACACAACTGCTCT
|-----|
102 107
112 CCTGTCACAGGTGGGAAAGAAACGAGCTGTGGGCCCTCTCTTTGGAT
...

12 matches found in sequence:
aaa16643 ; Human secreted protein clone qsl4_3 nucleotide sequence SEQ ID NO:5
(from "mycobacterng.seq")
TOIG of: aaa16643 check: 1366 from: 1 to: 5160
ID AAA16643 standard; cDNA; 5160 BP.
XX AC
XX AAA16643;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone qsl4_3 nucleotide sequence SEQ ID NO:51.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fachtel K;
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94922.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 60; Page 520-521; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in

```
217  GCCAGTCTCTAGCTGTGGATCCCAAAGGCGAGCCGTATGATTAG
...
1183  ATGGACTTGAGGTGTGAGAAATGGCTGTTCTGAGCTACCTGGTAACCCCAAGCGTGTCT
      |-----|
      1233 1238
1243  GGACAGTGGCTGCACACATTTGAAGATGAGTTTGATGCCCTACATCAT
...
1466  TGAGTGGAGACCCCTGGAAAGAAACAATTGTGAAGTGTGCAGTGAACCCAGCGACAAGT
      |-----|
      1516 1521
1526  GGTGATTGCCCTGCACAGGAGGAGAGTGCTGTATTTCGAGATGGAT
...
3388  ATGGAGACCTCTGTGAGCAGTTCAATTCCATGGAACCCCAACAACAAAGAAACGTCCTGTG
      |-----|
      3438 3443
3448  AAGAACTGGACCGAACCCCAACCCGGAAGTGTCCAAGAAACTCGAGGA
...
-----
2 matches found in sequence:
aaal6641 : Human secreted protein clone pm421_3 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6641 check: 3979 from: 1 to: 2635

ID  AAAL6641 standard; cDNA; 2635 BP.
XX  AC
XX  AAAL6641;
XX  DT 16-JUN-2000 (first entry)
XX  DE
XX  Human secreted protein clone pm421_3 nucleotide sequence SEQ ID NO:47.
KW  Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW  antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW  antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW  antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW  infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW  connective tissue disease; multiple sclerosis; erythematosis;
KW  rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW  Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW  insulin dependent diabetes mellitus; graft-versus-host-disease;
KW  autoimmune inflammatory eye disease; allergy; ss.
XX  OS Homo sapiens.
XX  PN WO200009552-A1.
XX  PD
XX  24-FEB-2000.
XX  PF 13-AUG-1999; 99WO-US18298.
XX  PR 14-AUG-1998; 98US-0096622.
XX  PR 17-AUG-1998; 98US-0096815.
XX  PR 04-SEP-1998; 98US-0099229.
XX  PR 23-OCT-1998; 98US-0105368.
XX  PR 08-JAN-1999; 99US-0115234.
XX  PR 12-FEB-1999; 99US-0119931.
XX  PR 18-FEB-1999; 99US-0120575.
XX  PR 30-APR-1999; 99US-0132020.
XX  PR 11-AUG-1999; 99US-0096622.
```

```
PA  (GEM ) GENETICS INST INC.
XX  Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI  Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI  Wong GG, Clark HF, Fectel K;
XX  WPI; 2000-205979/18.
DR  P-PSDB; AAY94921.
XX  New polynucleotides encoding secreted proteins, which may have e.g.
PT  nutritional, chemokine, immune stimulating or suppressing,
PT  hematopoiesis regulating, tissue growth, activin/inhibin
PT  antiinflammatory or tumor inhibition activity
XX  Claim 56; Page 517-518; 641pp; English.
XX  AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC  AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC  adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC  adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC  foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC  and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC  predicted to have biological activities which would make them suitable
CC  for treating, preventing or ameliorating medical conditions in humans
CC  and animals. The polynucleotides can be used as markers for tissues in
CC  which the protein is preferentially expressed, as molecular weight
CC  markers on Southern gels, and as chromosome markers or tags to identify
CC  chromosomes or to map gene positions. The proteins can be used in the
CC  treatment of immune deficiencies and disorders, such as severe combined
CC  immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC  infections. These infections include human immunodeficiency virus (HIV),
CC  hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC  candidiasis. The proteins can be used to treat autoimmune disorders such
CC  as connective tissue disease, multiple sclerosis, systemic lupus
CC  erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC  Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC  diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC  autoimmune inflammatory eye disease. The proteins can also be used to
CC  treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC  probes for the human secreted proteins from the present invention.
XX  Sequence 2635 BP; 702 A; 540 C; 502 G; 891 T; 0 other;
XX  AAAL6641 Length: 2635 March 5, 2002 14:18 Type: N Check: 3979
Found using 'seq2-3' (pappu403.key)
...
2379  TCCTTTTTTAAAGAAACACATGTAATACATTTAAGTATTGTAGGCATAGCGTTCAGA
      |-----|
      2452 2457
2439  TGTGACTGCCCGCGGTTCTCTGGACAAAGCCTGCATCCCGGTGATCAGCCACCTCA
      |-----|
      2452 2457
2499  AGCCCAAGG
...
-----
1 match found in sequence:
aaal6642 : Human secreted protein clone pv6_1 nucleotide sequence SEQ ID NO:49
(from "mycobacterng.seq")
TOIG of: aaal6642 check: 2777 from: 1 to: 1594
ID  AAAL6642 standard; cDNA; 1594 BP.
XX  AC
XX  AAAL6642;
XX  DT 16-JUN-2000 (first entry)
XX  DE Human secreted protein clone pv6_1 nucleotide sequence SEQ ID NO:49.
```

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 4433 BP; 1436 A; 746 C; 854 G; 1397 T; 0 other;

AAA16639 Length: 4433 March 5, 2002 14:18 Type: N Check: 409
Found using 'seq2-3' (pappu403.key)

...

43 GGCTCCGGGCCCTGCATCTCTAGCATGGACCGGACCTTTTGGCGCAGTCGCTAAAT
|-----|
93 98

103 TGCCACGGTGCCTTTTGTCTCTCTACTTCGGAGCGAACAGCAGGACATCCACACTTC
111 116

163 CGTA

...

981 ATCTTGGGCTTTTAAATGTGTGAAGATCGGAGTTGACTATTTTAAAGTTGGACGCCATGT
|-----|
1031 1036

1041 GGATGCTATGATGAATACAAATAAAGCTTTGGAAATAGACAAACA

...

3362 GAAGCAGGATAGGTTACAGTATGAAGAAGACACAGATAAAGAGAAAGATAGACGCCCTCT
|-----|
3412 3417

3422 CTCTTCATCTTCACTTGAATAACCGGATGATTTTGGAGTGCTACTCC

...

4211 CCATCTAGGTTGTGTAAGTACACTCTTATGATGTTTCACACAGTACAAATCGCCCAAG
|-----|
4261 4266

4271 GATGCATTATCAGAACACATTCCTCCATTGTTATCCAATGCATGACT

...

4 matches found in sequence:
aaal6640 : Human secreted protein clone pm412_12 nucleotide sequence SEQ ID NO
(from "mycobactereng.seq")
ToIG of: aaal6640 check: 2555 from: 1 to: 4017

ID AAA16640 standard; cDNA; 4017 BP.

XX AC

XX AC

XX AC

DT 16-JUN-2000 (first entry)

XX

DE

XX

Human secreted protein clone pm412_12 nucleotide sequence SEQ ID NO:45.
Human; secreted protein; immunestimulant; immunesuppressant; virucide;
antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
connective tissue disease; multiple sclerosis; erythematosis;

KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
OS Homo sapiens.
XX
XX
PN W0200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94920.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 54; Page 512-513; 64lpp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94988 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 4017 BP; 1023 A; 946 C; 1018 G; 1030 T; 0 other;

AAA16640 Length: 4017 March 5, 2002 14:18 Type: N Check: 2555
Found using 'seq2-3' (pappu403.key)

...

157 TGTGTGAGAAGATTACCAAGAACCTTTTGGCAAGAGTGGATGCATGCGCATCGTCTCTG

XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 50: Page 507-508; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAA94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1549 BP; 445 A; 269 C; 317 G; 518 T; 0 other;

AAA16638 Length: 1549 March 5, 2002 14:18 Type: N Check: 7151
Found using 'seq2-3' (pappu403.key)
...
417 TAATAAGAGCTGGGCTTGAGACTCGATCCACAGGCTTTGAGGCATTTGAAGGTCGTTGTT
467 472
477 TCAGATTGGCTGTAGTGCATGCTTATGGAGGCAAGTGCAGCTGC
...
717 GATTCAATACATGCTTGAGCATAGTCTTTTCCTCAGTGTGTATACTTAATAGCCGATAG
767 772
777 CCTCTATAGGAACGATGATGATTAGTCTGCTGCGCAGGAATGTTTT
...
1046 ATTTTTCACCACCTCTCTTTTGGTTAGTTGGAGCAGAGATCTGTTTCATCGCTTGAA
1096 1101
1106 TCAAAATATTGTTGGCATATCTGTTGCCACTCTAAGTTTGGCATAT
...
1378 TAGATGTCAGAATCTGCAGTGCCTCTTGAAAGAAGATCAAAATAGCATGCTGTTTCAT
1428 1433
1438 CCTTGAGCTTATCTCTTTGTTACTTTTACATTAAATCCITTTTAAAT
...

5 matches found in sequence:
aaa16639 ; Human secreted protein clone np26_3 nucleotide sequence SEQ ID NO:4
(from "mycobacterg.seq")
TOIG of: aaa16639 check: 409 from: 1 to: 4433

ID AAA16639 standard; cDNA; 4433 BP.
XX
AC AAA16639;
DT
XX 16-JUN-2000 (first entry)
DE Human secreted protein clone np26_3 nucleotide sequence SEQ ID NO:43.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antihyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94919.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 52; Page 509-510; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),

DE XX Human secreted protein clone cf85_1 nucleotide sequence SEQ ID NO:39.

KW XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;

KW XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;

KW XX antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;

KW XX antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

KW XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;

KW XX connective tissue disease; multiple sclerosis; erythematosis;

KW XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;

KW XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;

KW XX insulin dependent diabetes mellitus; graft-versus-host-disease;

KW XX autoimmune inflammatory eye disease; allergy; ss.

OS XX Homo sapiens.

XX XX

XX XX WO200009552-A1.

XX XX

XX XX 24-FEB-2000.

XX XX

XX XX 13-AUG-1999; 99WO-US18298.

XX XX

XX XX 14-AUG-1998; 98US-0096622.

XX XX 17-AUG-1998; 98US-0096815.

XX XX 04-SEP-1998; 98US-0099229.

XX XX 23-OCT-1998; 98US-0105368.

XX XX 08-JAN-1999; 99US-0115234.

XX XX 12-FEB-1999; 99US-0119931.

XX XX 18-FEB-1999; 99US-0120375.

XX XX 30-APR-1999; 99US-0132020.

XX XX 11-AUG-1999; 99US-0096622.

XX XX (GEMY) GENETICS INST INC.

XX XX

XX XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;

XX XX

DR WPI: 2000-205979/18.

XX P-PSDB; AAY94917.

XX XX

PT New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,

PT hematopoiesis regulating, tissue growth, activin/inhibin

PT antiinflammatory or tumor inhibition activity

XX XX

PS Claim 48; Page 506; 64lpp; English.

XX XX

CC AAA16618 to AAA16697 encode the human secreted proteins given in

CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,

CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,

CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,

CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,

CC and adult bladder, cDNA libraries. The polynucleotides and proteins are

CC predicted to have biological activities which would make them suitable

CC for treating, preventing or ameliorating medical conditions in humans

CC and animals. The polynucleotides can be used as markers for tissues in

CC which the protein is preferentially expressed, as molecular weight

CC markers on Southern gels, and as chromosome markers or tags to identify

CC chromosomes or to map gene positions. The proteins can be used in the

CC treatment of immune deficiencies and disorders, such as severe combined

CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such

CC as connective tissue disease, multiple sclerosis, systemic lupus

CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent

CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and

CC autoimmune inflammatory eye disease. The proteins can also be used to

CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent

XX probes for the human secreted proteins from the present invention.

XX XX

SQ Sequence 2015 BP; 562 A; 498 C; 388 G; 567 T; 0 other;

AAA16637 Length: 2015 March 5, 2002 14:18 Type: N Check: 8840
Found using 'seq2-3' (pappu403.key)

1131 TGATAGGCCTCGTCTCTCTGGGTAGATCTCTCGGAATCACTCCGACGAGGAACGTTACTC
1181 1186 |-----|

1191 AAGACTGGATTATTTGATCAATGGGATCTATGTGGACATCTAAGGA

1408 TGCAGTAGCACGATCTCGGCTCTCACCGCAACCTCCGTCCTCTGGTTCAGCGCATTTCTC
1458 1463 |-----|

1468 CTGCCTCAGCCTCTCTAGTATCTTGGGATTACAGGCATGTGCCACCAC

4 matches found in sequence:

aaa16638 ; Human secreted protein clone dd504_18 nucleotide sequence SEQ ID NO
(from 'mycobacterng.seq')
TOIG of: aaa16638 check: 7151 from: 1 to: 1549

ID AAA16638 standard; cDNA; 1549 BP.

XX XX
AC AAA16638;

XX XX 16-JUN-2000 (first entry)

XX XX Human secreted protein clone dd504_18 nucleotide sequence SEQ ID NO:41.

XX XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.

XX XX Homo sapiens.

XX XX OS

XX XX PN WO200009552-A1.

XX XX PD 24-FEB-2000.

XX XX PF 13-AUG-1999; 99WO-US18298.

XX XX PR 14-AUG-1998; 98US-0096622.

XX XX PR 17-AUG-1998; 98US-0096815.

XX XX PR 04-SEP-1998; 98US-0099229.

XX XX PR 23-OCT-1998; 98US-0105368.

XX XX PR 08-JAN-1999; 99US-0115234.

XX XX PR 12-FEB-1999; 99US-0119931.

XX XX PR 18-FEB-1999; 99US-0120375.

XX XX PR 30-APR-1999; 99US-0132020.

XX XX PR 11-AUG-1999; 99US-0096622.

XX XX (GEMY) GENETICS INST INC.

XX XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;

XX XX WPI: 2000-205979/18.

XX XX P-PSDB; AAY94918.

CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1759 BP; 586 A; 315 C; 300 G; 558 T; 0 other;

AAA16635 Length: 1759 March 5, 2002 14:18 Type: N Check: 5944 ..
Found using 'seq2-3' (pappu403.key)

...
76 GTATATAAATTCACAAAGTTAAGTGCACACAGTGTAGAGAGACATACAGCCTGAAC
126 131
136 TTCCATAACAGTCAATGGTACGTCAACATCATCATGTACAGAACA
...

3 matches found in sequence:
aaal6636 ; Human secreted protein clone rd810_6 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6636 check: 8795 from: 1 to: 643

ID AAA16636 standard; cDNA; 643 BP.
XX
AC AAA16636;
XX
DT 16-JUN-2000 (first entry)
DE Human secreted protein clone rd810_6 nucleotide sequence SEQ ID NO:37.
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fecthel K;
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94916.
XX
DR New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
PS Claim 46; Page 505; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 643 BP; 139 A; 190 C; 205 G; 109 T; 0 other;
AAA16636 Length: 643 March 5, 2002 14:18 Type: N Check: 8795 ..
Found using 'seq2-3' (pappu403.key)
...
72 TGGATGAGAAGTGAATCGGTCTCGGAGTCTCGAGTGACACGGCGGAGGCGGCCAGG
122 127
132 GACACTTTGGTTCTCCAGGGCTGGAAGGCTTCTAGAGGTTCTCATCAGGGAAGTGTG
192 GCTGGGGCGCGCTACTCTGGTGTACGACACGAGGAGTGCTGGGGCCCGCAAGAGG
198 203
252 CAGCGAGCCCTACAGAAGGTGGGGAGGTGTCCTCCCGCCGCCAT
...

2 matches found in sequence:
aaal6637 ; Human secreted protein clone cf85_1 nucleotide sequence SEQ ID NO:3
(from "mycobacterng.seq")
TOIG of: aaal6637 check: 8840 from: 1 to: 2015
ID AAA16637 standard; cDNA; 2015 BP.
XX
AC AAA16637;
XX
DT 16-JUN-2000 (first entry)
XX

PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94914.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 42; Page 503; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 988 BP; 324 A; 215 C; 227 G; 222 T; 0 other;

AAA16634 Length: 988 March 5, 2002 14:18 Type: N Check: 9644 ..
Found using 'seq2-3' (pappu403.key)
..
333 AGAATTTATGACTAAGTCATCAAAAGCATATGTATACAAAAAAGGCGCCGGG
|-----|
383 388
393 TGAGGGCGGAGCTGGGGCGATGCGTCGCGGAGCGCGCTCGCTAGTATTGGCACCCG
|-----|
415 420
453 TCAGGTCCGGGGCTCTCCGGAGCGGCCCTAGCTTGAGGAAGATGTCGTCCGCCG
|-----|
497 502
513 CATGGAGCGGCTCAGGCGGAGCGCTGGTACCGTCGAGGTGAGTCATCGTTACCCGCGAGG
|-----|
557 562 572
573 GCGCCATTTTGCCCAACCGGTGAAATGTCTTCCGCGCTTCTCCGTGTGTTCTC
577

...

1 match found in sequence:
aaal6635 : Human secreted protein clone rd610_1 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6635 check: 5944 from: 1 to: 1759
ID AAA16635 standard; cDNA; 1759 BP.
XX
AC AAA16635;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone rd610_1 nucleotide sequence SEQ ID NO:35.
XX
KW Human: secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI: 2000-205979/18.
DR P-PSDB; AAY94915.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 44; Page 504; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify

XX PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18298.
XX PR 14-AUG-1998; 98US-0096622.
XX PR 17-AUG-1998; 98US-0096815.
XX PR 04-SEP-1998; 98US-0099229.
XX PR 23-OCT-1998; 98US-0105368.
XX PR 08-JAN-1999; 99US-0115234.
XX PR 12-FEB-1999; 99US-0119931.
XX PR 18-FEB-1999; 99US-0120575.
XX PR 30-APR-1999; 99US-0132020.
XX PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
XX P-PSDB; AAY94913.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX Claim 40; Page 501-502; 64lpp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1527 BP; 403 A; 418 C; 386 G; 320 T; 0 other;

AAA16633 Length: 1527 March 5, 2002 14:18 Type: N Check: 7705
Found using 'seq2-3' (pappu403.key)

166 CATGGATGATACCTCAGGCTCTAGCTTCGAGGATATGGTGAGCTGCATCAGCGCTGGG |-----|
216 221
226 CGAGGAAGAAGTAGACGCTGATGCAGCTGATGCAGCTGCTGCTGAAGAGGAGGATGGAGA |-----|
239 244
286 GTTCCTGGG
...
518 CCTCTCATGCTGGTCTTTCACCTCTGGTTGCTATCTACTCTCCATGGGATGAAGACGCTGCAC |-----|
568 573
578 ACTATTATCCGGGAGGACCCCTGATGGGCACAGCCATTGGCACCT
...
630 GCTACTGGCTGGGAGTCTCATCTTCACTTACTTCTTGCCTACCTGTGCAACGCCACAGA |-----|
680 685
690 TCACCATGCTGCAGATGTTGGCACTGCTGGGCTATGGCCTCTTTGG
...
1011 GAGACATCCCTGCCCATCTCCCTGCTGCTGGGTTCCCAACACCGCTCCTCAACGCCACAG |-----|
1061 1066
1071 CCRAAGCTGTTCCGGTGACCTGCAGTCACACTGACCCCACTGNA
...

5 matches found in sequence:
aaa16634 : Human secreted protein clone pw337_6 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaa16634 check: 9644 from: 1 to: 988
ID AAA16634 standard; cDNA; 988 BP.
XX AC AAA16634;
XX AC AAA16634;
DT 16-JUN-2000 (first entry)
XX Human secreted protein clone pw337_6 nucleotide sequence SEQ ID NO:33.
XX Human: secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antinflammatory; dermatological;
KW antidiabetic; antisthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.

4 matches found in sequence:
aaal6632 : Human secreted protein clone pp314_19 nucleotide sequence SEQ ID NO
(from "mycobacterng.seq")
TOIG of: aaal6632 check: 7426 from: 1 to: 2472

ID AAAL6632 standard; cDNA; 2472 BP.
XX AAAL6632;
XX 16-JUN-2000 (first entry)
XX Human secreted protein clone pp314_19 nucleotide sequence SEQ ID NO:29.
DE Human; secreted protein; immunestimulant; immunesuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX Homo sapiens.
XX OS
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120375.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fachtel K;
XX
XX WPI; 2000-205979/18.
DR P-PSDB; AAY94912.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 38; Page 498-499; 641pp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined

immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2472 BP; 796 A; 475 C; 502 G; 699 T; 0 other;

AAAL6632 Length: 2472 March 5, 2002 14:18 Type: N Check: 7426 ..
Found using 'seq2-3' (pappu403.key)
...
72 ATTTGGGTGTACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAAAACAGCAGACA
122 127
125 130
132 TTCTGAGCAGATAGTCTGGGACTTTCAAAATCTTCAGAGAGCCAAAT
...
1857 GTGGCTGTACTGAAACTCATTTGGTGCATATATCAGACACAGACTCACAACGACCCAG
1907 1912
1917 AAGCATTCTCAAAGATGATGAATGTACTCAACAGAACTTCAATT
...
2128 ATTATACAAAGTCCAGGAGAACTCAACTGGTCTTTCTTAAGAAGCAGACACTGTCATTT
2178 2183
2188 TGTGCCACAAAGCTGCCTTCCACATGAAGCTGTGGGCTCAGCCAGC
...

8 matches found in sequence:
aaal6633 : Human secreted protein clone py35_1 nucleotide sequence SEQ ID NO:3
(from "mycobacterng.seq")
TOIG of: aaal6633 check: 7705 from: 1 to: 1527

ID AAAL6633 standard; cDNA; 1527 BP.
XX
XX AC AAAL6633;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone py35_1 nucleotide sequence SEQ ID NO:31.
XX
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX Homo sapiens.
XX OS
XX WO200009552-A1.

```
...
67 AGATAAGAGGAAGTTGCTTTATGGATGAATACTGTTGGCCCTACCATATCGTCAAGA
    |-----|
    117 122

127 AACATATAAGTACTACTTTTCACTTCCATCTGTTGTGGGTCAAAAAA
...

828 TATTTTCCTCTCTGATTGGTTCTGGATGTCAGATATTTGCTGTCTCTCATCGTTATTA
    |-----|
    878 883

888 TTGTTGCAATGATAGAGATTATATATACTGAGAGGGGATCAATGCT
...

1613 TACTTTGGATATATGCGGTATTAGCACAGCCTTGGGGATAATGTGTGAGCGGATGGT
    |-----|
    1663 1668

1673 TACATGGGAACAAGTGCCTTTGTTCGAAAAATCTATACTAATGTGA
...

-----
5 matches found in sequence:
aaal6631 ; Human secreted protein clone p174l_5 nucleotide sequence SEQ ID NO:
(from "mycobacterieng.seq")
TOIG of: aaal6631 check: 1935 from: 1 to: 3136

ID AAAL6631 standard; cDNA; 3136 BP.
XX AC AAAL6631;
XX DT 16-JUN-2000 (first entry)
XX DE Human secreted protein clone p174l_5 nucleotide sequence SEQ ID NO:27.
XX KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX OS Homo sapiens.
XX PN WO200009552-A1.
XX PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18298.
XX PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX PA (GEMY ) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie IA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
```

```
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSOB; AAY94911.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 36; Page 495-496; 641pp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC foetal placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 3136 BP; 892 A; 656 C; 688 G; 900 T; 0 other;
SQ
AAAL6631 Length: 3136 March 5, 2002 14:18 Type: N Check: 1935
Found using 'seq2-3' (pappu403.key)
1 GCTGCGAGTACCTCCATGCTGCCGGTGGCTGTGACGCGCGCAGTGCGCCTGTCTCTGCC
    |-----|
    45 50
61 ATAACAGCGCATTTTCTCAGATTTGCGGGAAATTAAGCAACTGCTGTATTGCGGGC
    |-----|
    67 72
121 CT
...
230 AGGAAGATGCCAGGATATGGATGCTTATACCTCTGGCCAAGGCTACTTTGACGTTAAAG
    |-----|
    280 285
290 AGTATGATCGGCGAGCAGCATTTCTTCTGATGGCTGCAATAGCAAGAA
...
338 CCTATTTTCTGTATATGTATTCCAGATATCTGTCTGGAGAAAAAAGACAGACGATGAAA
    |-----|
    388 393
398 CAGTTGATAGCTTAGGCCCCCTGGAAAAAGGACAAAGTGAAAAATGAGCGCTTAGAGAAAT
    |-----|
    444 449
458 TGAGAGTGGAGCTCAGCAAAAAACACCAAGCTCGAGAACTTG
...
```

CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2361 BP; 642 A; 502 C; 508 G; 709 T; 0 other;

AAA16629 Length: 2361 March 5, 2002 14:18 Type: N Check: 1246
Found using 'seq2-3' (pappu403.key)
...
38 CCTCATCCATGCCCTAACGCCAGAGTCCTGGCCTGTCGTAGTCAGTTTCAAGCGTTGGCT
88 93
98 CCTTGGGAGCCGATGAATCAAAAGTGTTATGTTCTGAGTTTAAAGA
...
1478 TTCTCTTTTTCACCGATGCCCTCTCTCAGCTTCTGAGTACGTCTCTGGGTGCGTGAG
1528 1533
1538 GTGATCCTAGGATCTGTCCTGAGACCAATGTGCTGTTTTCAGCCCC
...
2009 ATTTCCAGTTTGTATCAACCTCTTTATCTATATATTTAGGATAGAGTTCAAGCTTAGTC
2059 2064
2069 TTGAAAGATTTTCTAAAGTAGTCTTTCAAACTGTTCTCTCAGAGCCC
...
2116 AGGATTTTCCAAAAGTACCTTAGGAACCTTGTAGGCTGCGAGTGGGGTGTGGCGATAGAG
2166 2171
2176 CAGGAGGAGGAGACAGGGGCTGCAGGGGCTCCACCTCCCAACAG
...

3 matches found in sequence:
aaa16630 : Human secreted protein clone pk366_7 nucleotide sequence SEQ ID NO:
(from "mycobactering.seq")
TOIG of: aaa16630 check: 6606 from: 1 to: 3370

ID AAA16630 standard; cDNA; 3370 BP.
XX
AC AAA16630;
XX
DT
XX 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone pk366_7 nucleotide sequence SEQ ID NO:25.
XX
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;

KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antihyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
WO200009552-A1.
XX
24-FEB-2000.
XX
13-AUG-1999; 99WO-US18298.
XX
14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
(GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fecthel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94910.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
XX Claim 34; Page 493; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 3370 BP; 1031 A; 537 C; 676 G; 1126 T; 0 other;

AAA16630 Length: 3370 March 5, 2002 14:18 Type: N Check: 6606
Found using 'seq2-3' (pappu403.key)

PT antinflammatory or tumor inhibition activity -
XX Claim 30; Page 489-490; 641pp; English.
XX
CC AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependant
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 3349 BP; 1072 A; 603 C; 626 G; 1048 T; 0 other;
AAA16628 Length: 3349 March 5, 2002 14:18 Type: N Check: 5966 ..
Found using 'seq2-3' (pappu403.key)

2351 ACATTTAATAGGAGTACAATAAAACTGTGTCTAGCTTTTGTACAGAAAGCGCTAGAT
2401 2406
2411 ATTAAGAATTTTGAAATGGATCATCTTCTACTTCTGTGTCATTTTAA
...

4 matches found in sequence:
aaal6629 ; Human secreted protein clone mdl24_16 nucleotide sequence SEQ ID NO
TOIG of: aaal6629 check: 1246 from: 1 to: 2361
ID AAA16629 standard; cDNA; 2361 BP.
XX
AC AAA16629;
XX
DT 16-JUN-2000 (first entry)
XX
DE
XX
XX Human secreted protein clone mdl24_16 nucleotide sequence SEQ ID NO:23.
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependant diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB: AAY94909.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antinflammatory or tumor inhibition activity -
XX
PS Claim 32; Page 491-492; 641pp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable

CC probes for the human secreted proteins from the present invention.

XX Sequence 5263 BP; 1442 A; 1194 C; 1170 G; 1450 T; 7 other;

AA1627 Length: 5263 March 5, 2002 14:18 Type: N Check: 6142 ..
Found using "seq2-3" (pappu403.key)

```
1  AGTGAAGGAGCGCCTTGAGCTCGAGCGGCGCTGGCGGAGAGCGCGGCTGCTCC
   14 19 28 33 34 39
   |-----| |-----| |-----|
61  TCCCTCCCGCGGTATTATCTCTGGAGACACATCCACAGTTAGCACITTCCTCA
   |-----|
121 GATGCTGACGCTCGGTGAACAGTTGCCTTTGGTCACAAGATTTAGAACACACAGTGTCGA
   127 132
   |-----|
181 TC
...
1224 AGAAGATCCTGAAGCTCACAGAACNAATCAAGATTGCACAAACAGCCCGGAGACACAGC
   1274 1279 1280
   |-----| |-----| |---
-|
1284 TTGCTGAATCTTGAAGCTTGCCACAGTCGACAGAACACAGCGTGCCCG
   1285
...
1545 GTGTCAAAGTGGTTTTCAGCTTCTCCAGGCCACCCATTACAGCAGGCGCTGTAG
   1595 1600
   |-----|
1605 TCTCAAAGCCAGAGAGATTGCCTCACTCATTCGGAACAAATTTGG
...
1776 GTTCTAGTGCCACTTCAGGCTCAGTGGAGCCCAACAGACCAGGGGCGCATCGCTGTAG
   1826 1831
   |-----|
1836 GAGCATCCAGCTCCAAAACAAACACCCCTGGACATGCAGAGCTCAGG
...
1960 GGAACATTATCAGAGGACTATTCTCTTAATAATGCAGACTTACAGGAGGAGCGATAG
   2010 2015
   |-----|
2020 ATGTGACGATGGGAAGAACAGCTAAATGACCTTAACAGAGCTCCACCAGAATGAATCTT
   2025 2030
   |-----|
2080 G
...
2318 GNTCCACTGTAGCCAACTGTGTGTCCCTCATGAAGACTCGCAACAGGAGTTTCAGC
   2368 2373
   |-----|
2378 ACTTATTCTTGTTGTTTATTGCTTTCTCTGGAAGCACTGGGACGCGCTCTTCAGC
   2423 2428
   |-----|
```

```
2438 TATCTGGAACGGTCTTTTTCATCCCTAGATGATGCTGGCA
...
3630 AAAACCACTAGTAGGATAGGAGTATTTTTCCTCCCAAGAAACGTTTCACC
   |-----|
   3680 3685
3690 CAGTGATCTTGGGCTGGGTTGTCTTTAGGAAAAAGTTGAGACTATA
...
5171 GAGCTCTTCTATAACCTCGCTCTTAATGCTTAACAGTTGTTCTGGTGGAAACGTCGCCCT
   |-----|
   5221 5226
5231 CATTTGAATGCTCCTCTAAAAAAA
-----
6 matches found in sequence:
aa16628 ; Human secreted protein clone ci52_2 nucleotide sequence SEQ ID NO:2
(from "mycobacterng.seq")
TOIG of: aa16628 check: 5966 from: 1 to: 3349

ID AA16628 standard; cDNA; 3349 BP.
XX
AC AA16628;
XX
DT
XX
XX 16-JUN-2000 (first entry)
DE
XX Human secreted protein clone ci52_2 nucleotide sequence SEQ ID NO:21.
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-009229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
PA (GEMY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
DR WPI; 2000-205979/18.
DR P-PSDB; AAY94908.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
```



```
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 1264 BP; 440 A; 247 C; 240 G; 337 T; 0 other;
..
AAA16624 Length: 1264 March 5, 2002 14:18 Type: N Check: 5763 ..
Found using 'seq2-3' (pappu403.key)
...
34 AACCACTAGCAAGAAGATCTTTAAAGGAGTTTGGTAGCCGAACTGTAGGCGCTTTTG
      84 89
      |-----|
94 GAGCATATTTTTTTTGTAGCAAGATGCACACAAAGCAAGATTTCAG
...
485 AGTGCAGTGGCGTAATCTTGGTCCACGGCAACCTCTGACTCCTGGATTTCAGGATTCCTC
      535 540
      |-----|
545 GTGCCTCTGCTCTCTCGAGTAGCTGGGACTACAGGTGGTGCCACCA
...
676 ATCTCTCTGCTGGCGCTGCCGAGTGCTGGGATTACAGGTGTGAGCCACAGCGCCTGGC
      726 731
      |-----|
736 CCCAAATATTTCTTAATCTTCCACTGTGANTTGCATGATATTCCTTA
...
-----
4 matches found in sequence:
aaal6625 ; Human secreted protein clone qg596_12 nucleotide sequence SEQ ID NO
(from "mycobacterng.seq")
TOIG of: aaal6625 check: 1053 from: 1 to: 2671
ID AAA16625 standard; cDNA; 2671 BP.
XX
AC AAA16625;
XX
DT 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone qg596_12 nucleotide sequence SEQ ID NO:15.
DE
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
OS Homo sapiens.
XX
PN WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
```

```
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY ) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94905.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
PS Claim 24; Page 478-479; 64lpp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
XX probes for the human secreted proteins from the present invention.
XX
SQ Sequence 2671 BP; 787 A; 567 C; 686 G; 631 T; 0 other;
..
AAA16625 Length: 2671 March 5, 2002 14:18 Type: N Check: 1053 ..
Found using 'seq2-3' (pappu403.key)
...
608 ACTGAAGAAACAATAAGATACCAAGGAGAGAGCTAAGTAAGTAAGCAGTGGAGAGCGCTTCC
      658 663
      |-----|
668 GCATGACAAACTCCTCTCTTGTAGACACTCAACAGGAGGAGGAGGATG
...
841 TTTCAGGCTATGTTTCGAGGCGAGACTCTGAATGTCAATAGTGTCTGCATATATCGTTGGAT
      891 896
      |-----|
901 ATGCTGATTTGCCAGATGAACAGATAGATGCCCGGAGAGCCCTTT
...
|-----|
```

PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX Claim 20; Page 474-475; 641pp; English.
XX
CC AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 3262 BP; 794 A; 859 C; 894 G; 715 T; 0 other;

AAAL6623 Length: 3262 March 5, 2002 14:18 Type: N Check: 6905
Found using 'seq2-3' (pappu403.key)
...
1138 TCTCGAGNAGGAGGAGGAGATGATGATGACACAGTCTGGAGGGGAGACGCTTCCCG
1188 1193
1198 CTGGAACGGGATGAAGTGATGCTCTCCCATCTACAGAGGTGGCGCATCGGCTGAACGACGACT
...
1595 CTGCAGCACCACCTCAAAAGSCCAACAGACTCAATCAACATCCTAGCGGACGCTTGTG
1645 1650
1655 CTGAGGAGATCCCAACACAGTGTTCGAGAGCATGAAGCTGGGGGT
...
2198 GCAACATGACAGACCATGTCTCCCATCTACAGAGGTGGCGCATCGGCTGAACGACGACT
2248 2253
2258 GGGCATACGGCAATGATCTTGTATGCCCGGCTTGGGACTTCCAGGACAGGAGTGTGCC
2318 TCTGTCGAACATGAACGCTTCAACGCCCGCGCTATGACCGGGCCACAGCAACCGTG
2333 2338 2346 2353
2378 ACTTCTGCCAGTGGACAACCTGCTG
...

3 matches found in sequence:
aaal6624 : Human secreted protein clone qc297_15 nucleotide sequence SEQ ID NO
(from "mycobacter.ng.seq")

TOIG of: aaal6624 check: 5763 from: 1 to: 1264

AAAL6624 standard; cDNA; 1264 BP.

AAAL6624;

16-JUN-2000 (first entry)

Human secreted protein clone qc297_15 nucleotide sequence SEQ ID NO:13.
Human; secreted protein; immunestimulant; immunosuppressant; virucide;
antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
antitubercular; immune deficiency; severe combined immunodeficiency; SCID;
infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
connective tissue disease; multiple sclerosis; erythematosus;
rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
insulin dependent diabetes mellitus; graft-versus-host-disease;
autoimmune inflammatory eye disease; allergy; ss.

Homo sapiens.

WO200009552-A1.

24-FEB-2000.

13-AUG-1999; 99WO-US18298.

14-AUG-1998; 98US-0096622.

17-AUG-1998; 98US-0096815.

04-SEP-1998; 98US-0099229.

23-OCT-1998; 98US-0105368.

08-JAN-1999; 99US-0115234.

12-FEB-1999; 99US-0119931.

18-FEB-1999; 99US-0120575.

30-APR-1999; 99US-0132020.

11-AUG-1999; 99US-0096622.

(GEMY) GENETICS INST INC.

Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;

Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

Wong GG, Clark HF, Fehel K;

WPI; 2000-205979/18.

P-PSDB; AAY94904.

New polynucleotides encoding secreted proteins, which may have e.g.

nutritional, chemokine, immune stimulating or suppressing,

hematopoiesis regulating, tissue growth, activin/inhibin

antiinflammatory or tumor inhibition activity

Claim 22; Page 477-478; 641pp; English.

AAAL6618 to AAAL6697 encode the human secreted proteins given in

AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,

adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,

adult placenta, adult testis, whole embryo, adult cartilage, kidney,

foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,

and adult bladder, cDNA libraries. The polynucleotides and proteins are

predicted to have biological activities which would make them suitable

for treating, preventing or ameliorating medical conditions in humans

and animals. The polynucleotides can be used as markers for tissues in

CC which the protein is preferentially expressed, as molecular weight

CC markers on Southern gels, and as chromosome markers or tags to identify

CC chromosomes or to map gene positions. The proteins can be used in the

CC treatment of immune deficiencies and disorders, such as severe combined

CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other

CC infections. These infections include human immunodeficiency virus (HIV),

CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such

CC as connective tissue disease, multiple sclerosis, systemic lupus

PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18298.
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94903.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX Claim 18; Page 472; 641pp; English.
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAAL6618 to AAAL6697, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX SQ Sequence 2104 BP: 695 A; 373 C; 391 G; 644 T; 1 other;
AAAL6622 Length: 2104 March 5, 2002 14:18 Type: N Check: 2766
Found using 'seq2-3' (pappu403.key)

19 ACAGTCTCTCGGNAATCAGTCTGTGATTTTCGGGAGACACAGACGGGACGGCG
69 74 75
|-----|
|-----|

79 CTCTTGTGGTGCATCTGGGCCAGGTGACGAGAACAGTTTCTCTGTTGAG
80
-|

...
152 CCCACACCCCTAGGCCCTAAAGATGCTGAGGTCTGTATGGAATTTCTGGAACGCCACAA
202 207
|-----|
212 AAAGAAATGCATCTTCTCTGGGCACGGTCTTGGAGGAGTATATATT
...
281 AATCAGAGAAATACAGGAAAGGAGGCTGCGAATACATTGCCCAAGCAGCAGCAATA
331 336
|-----|
341 TCATTTGAAAGTAACACGAGGAGCTTGCATATATACACAGTCTGCTCC
...

7 matches found in sequence:
aaal6623 ; Human secreted protein clone pt332_1 nucleotide sequence SEQ ID NO: 1
(from 'mycobacterieng.seq')
TOIG of: aaal6623 check: 6905 from: 1 to: 3262
ID AAAL6623 standard; cDNA; 3262 BP.
XX AC AAAL6623;
XX DT 16-JUN-2000 (first entry)
XX DE Human secreted protein clone pt332_1 nucleotide sequence SEQ ID NO:11.
XX KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antisthmatic; antiarthritic; antirheumatic; protozoacide;
KW antihydroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosus;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
OS Homo sapiens.
XX PN WO200009552-A1.
XX PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18298.
XX PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEMY) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94903.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,

aaal6621 : Human secreted protein clone pj193_5 nucleotide sequence SEQ ID NO: (from "mycobacterng.seq")
TOIG of: aaal6621 check: 1132 from: 1 to: 1462

ID AAAL6621 standard; cDNA; 1462 BP.
XX AC AAAL6621;
XX DT 16-JUN-2000 (first entry)
XX DE Human secreted protein clone pj193_5 nucleotide sequence SEQ ID NO:7.
XX KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX OS Homo sapiens.
XX PN WO200009552-A1.
XX PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18298.
XX PR 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX PA (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, LaVallie BR, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94901.
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity .
XX Claim 16; Page 470-471; 641pp; English.
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder. cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and

CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependant
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
CC probes for the human secreted proteins from the present invention.
XX SQ Sequence 1462 BP; 384 A; 316 C; 431 G; 331 T; 0 other;
AAAL6621 Length: 1462 March 5, 2002 14:18 Type: N Check: 1132
Found using 'seq2-3' (pappu403.key)
...
25 ACCCACACTACCTCTCCCGAAGTTGAAGCAAGCGGTGATTGTTGTAGACGCGCTTTGT
|-----|
75 80
85 CATGGGACCTGTGCGGTGGGAATATTGCTTTTCTTTTGGCCGTCACGAGGCGTGT
|-----|
145 GGCTGGGATGTTGAAGGAGGAGGACGATGACACAGAACGCTTGCCACGAAATCGGAAGT
167 172 180 185
205 GTGTAAGCTGCTGAGCACAGACTACAGCGG
|-----|
...
273 AGCTGGGCGAGGTCTGGATACAGGCAAGGAAGAGACACAGCGTCCCTACAGCCTTCAG
|-----|
323 328
333 AGACAGGCTGGAAGAGGCGCTTAGAGAATTTATGTGAGCGGATCCT
|-----|
...
1369 AATGTAACCCAGAGCCACCATATAGTTTATAGTGCTCAATTTTCTATATCCTATTATA
|-----|
1419 1424
1429 AACTTTTTTCTTTTCTAAAAAATAAAAAA
|-----|
4 matches found in sequence:
aaal6622 : Human secreted protein clone pj317_2 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6622 check: 2766 from: 1 to: 2104

ID AAAL6622 standard; cDNA; 2104 BP.
XX AC AAAL6622;
XX DT 16-JUN-2000 (first entry)
XX DE Human secreted protein clone pj317_2 nucleotide sequence SEQ ID NO:9.
XX KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX OS Homo sapiens.
XX PN WO200009552-A1.
XX


```
61 GCAGAGT
...
614 ATGCCAAAGACGAGACTACCTGCTTCTGGTGATGGAGGCGACAGATGACGGCGCCTGCT
      |-----|
      564 569

674 GGAGTATGATACTGTGACCAAGGAAGTAAAGTTTATTTGACACG
...
813 AAGAGACGGTGATGAAGTTTTCGCCGGTACACGCTCGTCTAGAACCTCAGCGACAGCG
      |-----|
      863 868

873 GTGCCCTCCGGAGAACCTGCATGATCCGATGGCTGGTGCCAC
...
1664 GGAATTGGATGATTCTGCATTCGGGACCTACTGCAGGGGTCCGTTTAGTAACGCTTGTG
      |-----|
      1714 1719

1724 CTGTGATCTTTGTTCTTGACCTCTAGACCCCAAGATGTGAACAGTG
...
1819 TATATTTTCAAGTATGCTCTACACACTGGTTCATGATTTTGATATAAATAACGATAAAT
      |-----|
      1869 1874

1879 CGACTTCTGCTGATTACCTTTAAAAAATAAAAAAAAAAAAAAAAAA
...
-----
9 matches found in sequence:
aaal6619 : Human secreted protein clone lo311_8 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6619 check: 5593 from: 1 to: 3508

ID AAA16619 standard; cDNA; 3508 BP.
XX
AC
AC
XX
DT 16-JUN-2000 (first entry)
XX
XX
DE
XX
KW Human secreted protein clone lo311_8 nucleotide sequence SEQ ID NO:3.
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX
XX 17-AUG-1998; 98US-0096815.
XX
XX 04-SEP-1998; 98US-0099229.
XX
XX 23-OCT-1998; 98US-0105368.
```

```
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GEM ) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Rechtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB: AAY94899.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 12: Page 467-468; 641pp; English.
PS
XX AAA16618 to AAA16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC and animals. The polynucleotides can be used as markers for tissues in
CC which the protein is preferentially expressed, as molecular weight
CC markers on Southern gels, and as chromosome markers or tags to identify
CC chromosomes or to map gene positions. The proteins can be used in the
CC treatment of immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
XX Sequence 3508 BP; 1094 A; 609 C; 697 G; 1108 T; 0 other;
SQ
...
AAA16619 Length: 3508 March 5, 2002 14:18 Type: N Check: 5593
Found using 'seq2-3' (pappu403.key)
...
6 AACATGCCAGCGGTTGGCGGTTTGGTGTGCTCTCTGTGACCATTGGTGGTGGCCCTGCTC
      |-----| |-----|
      56 71 75 80 56 61
66 ATCGTTTGGCAGCTGTCCCTCAGCTCTGCCCCAAGAAAGAGAGAGATGGTGTATTATCTGAA
      |-----|
      66 71 75 80
126 AAGGT
...
143 GGAATGGACTAACAAAAGACCTGTATAAGAAATGAATGAGACAAAGTCCGTCGCCCTTGT
      |-----|
      193 198
203 GAAAGCCCCACCGAGAAATTAATCCCGTTACGTCATGTTCTCACTGCTCTCCAACTGCATAG
      |-----|
      231 236
```

```
> O <
O I 10 IntelliGenetics
> O <

Quest - Quick User-directed Expression Search Tool
Release 5.4

-- Outline of search "seq2-3varng" --

Selected search type is key against sequence data banks or files.
Selected scope is Sequence.
Selected sequence key from "pappu403.key":
seq2-3 (NA) ID seq2-3 NA preliminary pattern
1 either-or
2 followed by
3 a or g
4 a or g
4 cg
4 c or t
4 c or t
4 c or t
4 followed by
4 a or g
4 tcg
4 c or t
4 c or t
2 followed by
3 a or g
3 a or g
3 cga
3 c or t

Selected files:
File : mycobacterng.seq

-- Output Parameters --

Format Options:
Nucleic acid code matching Exact Indirect file NO
Find non-matching hits only No Sequence or key file NO
Report key used Yes List of hits Yes
Note position of hit Yes Hit display Yes
Display full annotations Yes Name and annotations Yes
Sequence context 50

Run mode Batch
Time to start comparison now
Notify at end of run NO

-----
6 matches found in sequence:
aaal6618 : Human secreted protein clone cof2_12 nucleotide sequence SEQ ID NO:
(from "mycobacterng.seq")
TOIG of: aaal6618 check: 3688 from: 1 to: 1925

ID AAAL6618 standard; cDNA: 1925 BP.
XX
AC AAAL6618;
XX
DT 16-JUN-2000 (first entry)
XX
DE Human secreted protein clone cof2_12 nucleotide sequence SEQ ID NO:1.
XX
KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antitumoric; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
```

```
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
OS Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fechtel K;
XX
XX WPI; 2000-205979/18.
XX P-PSDB; AAY94898.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity
XX
XX Claim 1; Page 465-466; 641pp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
XX AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX which the protein is preferentially expressed, as molecular weight
XX markers on Southern gels, and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX candidiasis. The proteins can be used to treat autoimmune disorders such
XX as connective tissue disease, multiple sclerosis, systemic lupus
XX erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
XX Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
XX diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
XX autoimmune inflammatory eye disease. The proteins can also be used to
XX treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
XX probes for the human secreted proteins from the present invention.
XX
XX Sequence 1925 BP; 460 A; 454 C; 520 G; 491 T; 0 other;
SQ
```

```
AAAL6618 Length: 1925 March 5, 2002 14:18 Type: N Check: 3688
Found using 'seq2-3' (pappu403.key)
|-----|
|-----|
3 8 12 17
1 AGTCGTCTACAGACGATGATGCCAGCCCGGAGGCTAAGAGCGCAGCTCCTTTAGCG
```

> 0 <
 01 10 IntelliGenetics
 > 0 <

Quest - Quick User-directed Expression Search Tool
 Release 5.4

-- Outline of search "seq4ng" --

Selected search type is key against sequence data banks or files.
 Selected scope is Sequence.

Selected sequence key from "pappu403.key":

1 seq4 (NA) ID seq4 NA preliminary pattern
 2 either-or
 2 aacgttcg
 2 cgaacgtt

Selected files:

File : mycobacterng.seq

-- Output Parameters --

| Format Options: | | File Options: | | |
|-----------------------------|-------|----------------------|-----|--|
| Nucleic acid code matching | Exact | Indirect file | No | |
| Find non-matching hits only | No | Sequence or key file | No | |
| Report key used | Yes | List of hits | Yes | |
| Note position of hit | Yes | Hit display | Yes | |
| Display full annotations | Yes | Name and annotations | Yes | |
| Sequence context | 50 | | | |

-- Run Parameters --

| Run mode | Batch |
|--------------------------|-------|
| Time to start comparison | now |
| Notify at end of run | No |

1 match found in sequence:

aah73439 ; Immunomodulatory nucleic acid.
 (from "mycobacterng.seq")
 TOIG of: aah73439 check: 8143 from: 1 to: 22

ID AAH73439 standard; DNA: 22 BP.

XX AC AAH73439;

XX DT 01-OCT-2001 (first entry)

XX DE Immunomodulatory nucleic acid.

XX KW G3PDH gene; immunomodulatory oligonucleotide; infection; mycobacterium;
 XX KW intracellular pathogen; anti-pathogenic; ss.

XX OS Unidentified.

XX PN WO200155341-A2.

XX PD 02-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US03029.

XX PR 31-JAN-2000; 2000US-0179353.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Raz E, Kornbluth R, Catanzaro A, Hayashi T, Carson DA;

XX DR WPI; 2001-483234/52.

XX PT Treating infection of intracellular pathogen e.g., Mycobacterium, in a
 PT subject, involves administering immunomodulatory nucleic acid molecule

PT to inhibit intracellular replication of intracellular pathogen -
 XX Examples; Page 26; 54pp; English.

XX The present invention describes a method of treating an infection caused
 CC by an intracellular pathogen, involving administering to the patient an
 CC immunomodulatory nucleic acid and an anti-pathogenic agent. This is
 CC particularly useful in the treatment of mycobacterial infections. The
 CC present sequence is an immunomodulatory nucleic acid described in the
 CC exemplification of the invention.

XX Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;

AAH73439 Length: 22 March 5, 2002 14:18 Type: N Check: 8143

Found using 'seq4' (pappu403.key)

1 TGACTGTGACGTCGAGATGA
 9
 16

1 match found in sequence:
 aav80096 ; Immunomodulatory oligo comprising an ISS sequence.

(from "mycobacterng.seq")
 TOIG of: aav80096 check: 8058 from: 1 to: 22

ID AAV80096 standard; DNA: 22 BP.

XX AC AAV80096;

XX DT 12-MAR-1999 (first entry)

XX DE Immunomodulatory oligo comprising an ISS sequence.

XX KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.

OS Synthetic.

PN WO9855495-A2.

XX PD 10-DEC-1998.

XX PF 05-JUN-1998; 98WO-US11578.

XX PR 06-JUN-1997; 97US-0048793.

XX PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.

XX PI Dina D, Roman M, Schwartz D;

XX DR WPI; 1999-059898/05.

XX Immunostimulatory oligonucleotides regulate the immune system - and
 PT contain an immune-stimulating octanucleotide sequence; for treating
 PT cancer, allergic and infectious diseases

XX Claim 7; Page 29; 63pp; English.

XX The invention relates to immunomodulatory oligonucleotides that comprise
 CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
 CC sequences are selected from the group consisting of AACGTCC, AACGTTCG,
 CC GACGTCC, and GACGTTCG. The immunomodulatory sequences are used to treat
 CC patients needing immune regulation, such as those suffering from cancer,
 CC an allergic disease and asthma. They are also used to prevent infectious
 CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
 CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
 CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
 CC Schistosoma. The immunomodulatory sequences are used to screen for human
 CC immunostimulatory activity by incubating macrophage cells and the
 CC oligonucleotide; and determining the relative amount of Th1-biased

CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
 CC specific claimed examples of such immunomodulatory oligonucleotides.
 XX
 SQ Sequence 22 BP; 6 A; 4 C; 7 G; 5 T; 0 other;

AAV80096 Length: 22 March 5, 2002 14:19 Type: N Check: 8058
 Found using 'seq4' (pappu403.key)

1 TGACCGTGAACGTTTCGAGATGA
 9 16

1 match found in sequence:
 aav80097; Immunomodulatory oligo comprising an ISS sequence.
 (from "mycobacterng.seq")
 TOIG of: aav80097 check: 8143 from: 1 to: 22

ID AAV80097 standard; DNA; 22 BP.
 XX
 AC AAV80097;
 XX
 DT 12-MAR-1999 (first entry)
 XX
 DE Immunomodulatory oligo comprising an ISS sequence.
 XX
 KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX
 OS Synthetic.

XX WO9855495-A2.
 XX 10-DEC-1998.

XX 05-JUN-1998; 98WO-US11578.

XX 06-JUN-1997; 97US-0048793.

XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.

XX Dina D, Roman M, Schwartz D;

XX WPI; 1999-059898/05.

XX Immunostimulatory oligonucleotides regulate the immune system - and
 XX contain an immune-stimulating octanucleotide sequence; for treating
 XX cancer, allergic and infectious diseases

XX Claim 5; Page 29; 63pp; English.

XX The invention relates to immunomodulatory oligonucleotides that comprise
 XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
 XX sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
 XX GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
 XX patients needing immune regulation, such as those suffering from cancer,
 XX an allergic disease and asthma. They are also used to prevent infectious
 XX diseases such as influenza, herpes, hepatitis B, human immunodeficiency
 XX and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
 XX Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
 XX Schistosoma. The immunomodulatory sequences are used to screen for human
 XX immunostimulatory activity by incubating macrophage cells and the
 XX oligonucleotide; and determining the relative amount of Th1-biased
 XX cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
 XX specific claimed examples of such immunomodulatory oligonucleotides.

XX Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;

AAV80097 Length: 22 March 5, 2002 14:19 Type: N Check: 8143

Found using 'seq4' (pappu403.key)

1 TGACTGTGAACGTTTCGAGATGA
 9 16

1 match found in sequence:
 aav80098; Immunomodulatory oligo comprising an ISS sequence.
 (from "mycobacterng.seq")
 TOIG of: aav80098 check: 9493 from: 1 to: 23

ID AAV80098 standard; DNA; 23 BP.
 XX
 AC AAV80098;
 XX
 DT 12-MAR-1999 (first entry)
 XX
 DE Immunomodulatory oligo comprising an ISS sequence.

XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX
 OS Synthetic.

XX WO9855495-A2.

XX 10-DEC-1998.

XX 05-JUN-1998; 98WO-US11578.

XX 06-JUN-1997; 97US-0048793.

XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.

XX Dina D, Roman M, Schwartz D;

XX WPI; 1999-059898/05.

XX Immunostimulatory oligonucleotides regulate the immune system - and
 XX contain an immune-stimulating octanucleotide sequence; for treating
 XX cancer, allergic and infectious diseases

XX Claim 6; Page 29; 63pp; English.

XX The invention relates to immunomodulatory oligonucleotides that comprise
 XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
 XX sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
 XX GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
 XX patients needing immune regulation, such as those suffering from cancer,
 XX an allergic disease and asthma. They are also used to prevent infectious
 XX diseases such as influenza, herpes, hepatitis B, human immunodeficiency
 XX and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
 XX Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
 XX Schistosoma. The immunomodulatory sequences are used to screen for human
 XX immunostimulatory activity by incubating macrophage cells and the
 XX oligonucleotide; and determining the relative amount of Th1-biased
 XX cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
 XX specific claimed examples of such immunomodulatory oligonucleotides.

XX Sequence 23 BP; 6 A; 8 C; 3 G; 6 T; 0 other;

AAV80098 Length: 23 March 5, 2002 14:19 Type: N Check: 9493

Found using 'seq4' (pappu403.key)

1 TCATCTCGAACGTTCCACACTCA
 7 14

2 matches found in sequence:

aav80100; Immunomodulatory oligo comprising an ISS sequence.
 (from "mycobacterng.seq")

TOIG of: aav80100 check: 5564 from: 1 to: 26

ID AAV80100 standard; DNA; 26 BP.
 AC AAV80100;
 DT 12-MAR-1999 (first entry)
 XX
 XX Immunomodulatory oligo comprising an ISS sequence.
 DE
 XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX
 OS Synthetic.
 XX
 PN WO9855495-A2.
 XX
 PD 10-DEC-1998.
 XX
 PF 05-JUN-1998; 98WO-US11578.
 XX
 PR 06-JUN-1997; 97US-0048793.
 XX
 PA (DYNA-) DYNAX TECHNOLOGIES CORP.
 XX
 PI Dina D, Roman M, Schwartz D;
 XX WPI; 1999-059898/05.
 DR
 XX Immunostimulatory oligonucleotides regulate the immune system - and
 PT contain an immune-stimulating octanucleotide sequence; for treating
 PT cancer, allergic and infectious diseases
 XX
 PS Claim 9; Page 29; 63pp; English.
 XX

The invention relates to immunomodulatory oligonucleotides that comprise at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS sequences are selected from the group consisting of AACGTTC, AACGTTCG, GACGTTC, and GACGTTCG. The immunomodulatory sequences are used to treat patients needing immune regulation, such as those suffering from cancer, an allergic disease and asthma. They are also used to prevent infectious diseases such as influenza, herpes, hepatitis B, human immunodeficiency and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and Schistosoma. The immunomodulatory sequences are used to screen for human immunostimulatory activity by incubating macrophage cells and the oligonucleotide; and determining the relative amount of Th1-biased cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent specific claimed examples of such immunomodulatory oligonucleotides.

Sequence 26 BP; 5 A; 9 C; 4 G; 8 T; 0 other;

AAV80100 Length: 26 March 5, 2002 14:19 Type: N Check: 5564
 Found using 'seq4' (pappu403.key)

1 TCCATAACGTCGCCATAACGTCGTC
 6 13 17 24
 |-----| |-----|

1 match found in sequence:
 aav80102; Immunomodulatory oligo comprising an ISS sequence.
 (from "mycobactering.seq")
 TOIG of: aav80102 check: 8143 from: 1 to: 22

ID AAV80102 standard; DNA; 22 BP.
 XX
 AC AAV80102;
 XX
 DT 12-MAR-1999 (first entry)
 XX

Immunomodulatory oligo comprising an ISS sequence.

DE
 XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX
 OS Synthetic.
 XX

Key Location/Qualifiers
 modified_base 11
 /*tag= a
 /note= "5-bromocytosine"
 XX

WO9855495-A2.

10-DEC-1998.

05-JUN-1998; 98WO-US11578.

06-JUN-1997; 97US-0048793.

(DYNA-) DYNAX TECHNOLOGIES CORP.

Dina D, Roman M, Schwartz D;

WPI; 1999-059898/05.

Immunostimulatory oligonucleotides regulate the immune system - and contain an immune-stimulating octanucleotide sequence; for treating cancer, allergic and infectious diseases

Claim 23; Page 30; 63pp; English.

The invention relates to immunomodulatory oligonucleotides that comprise at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS sequences are selected from the group consisting of AACGTTC, AACGTTCG, GACGTTC, and GACGTTCG. The immunomodulatory sequences are used to treat patients needing immune regulation, such as those suffering from cancer, an allergic disease and asthma. They are also used to prevent infectious diseases such as influenza, herpes, hepatitis B, human immunodeficiency and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and Schistosoma. The immunomodulatory sequences are used to screen for human immunostimulatory activity by incubating macrophage cells and the oligonucleotide; and determining the relative amount of Th1-biased cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent specific claimed examples of such immunomodulatory oligonucleotides.

Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;

AAV80102 Length: 22 March 5, 2002 14:19 Type: N Check: 8143
 Found using 'seq4' (pappu403.key)

1 TGACTGTGACGTCGAGATGA
 9 16
 |-----|

1 match found in sequence:
 aav80103; Immunomodulatory oligo comprising an ISS sequence.
 (from "mycobactering.seq")
 TOIG of: aav80103 check: 8143 from: 1 to: 22

ID AAV80103 standard; DNA; 22 BP.
 XX
 AC AAV80103;
 XX
 DT 12-MAR-1999 (first entry)
 XX

Immunomodulatory oligo comprising an ISS sequence.
 Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 XX

KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX Synthetic.

XX Key Location/Qualifiers
 XX modified_base 11
 FT /*tag= a
 FT /note= "5-bromocytosine"

XX WO9855495-A2.

XX 10-DEC-1998.

XX 05-JUN-1998; 98WO-US11578.

XX 06-JUN-1997; 97US-0048793.

XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.

XX Dina D, Roman M, Schwartz D;

XX WPI; 1999-059898/05.

XX Immunostimulatory oligonucleotides regulate the immune system - and
 PT contain an immune-stimulating octanucleotide sequence; for treating
 PT cancer, allergic and infectious diseases

XX Claim 24; Page 30; 63pp; English.

XX The invention relates to immunomodulatory oligonucleotides that comprise
 CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
 CC sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
 CC GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
 CC patients needing immune regulation, such as those suffering from cancer,
 CC an allergic disease and asthma. They are also used to prevent infectious
 CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
 CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
 CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
 CC Schistosoma. The immunomodulatory sequences are used to screen for human
 CC immunostimulatory activity by incubating macrophage cells and the
 CC oligonucleotide; and determining the relative amount of Th1-biased
 CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
 CC specific claimed examples of such immunomodulatory oligonucleotides.

XX Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;

AAV80103 Length: 22 March 5, 2002 14:19 Type: N Check: 8143

Found using 'seq4' (pappu403.key)

1 TGACTGTGACGTTTCAGATGA
 9 16

1 match found in sequence:

aaV80105 ; Oligo used in experiments for stimulation of cytokine production.
 (from "mycobacterng.seq")
 TOIG of: aav80105 check: 8045 from: 1 to: 22

ID AAV80105 standard; DNA; 22 BP.

XX AAV80105;

XX 12-MAR-1999 (first entry)

XX Oligo used in experiments for stimulation of cytokine production.

XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.

XX Synthetic.
 XX WO9855495-A2.
 XX 10-DEC-1998.
 XX 05-JUN-1998; 98WO-US11578.
 XX 06-JUN-1997; 97US-0048793.
 XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
 XX Dina D, Roman M, Schwartz D;
 XX WPI; 1999-059898/05.

XX Immunostimulatory oligonucleotides regulate the immune system - and
 PT contain an immune-stimulating octanucleotide sequence; for treating
 PT cancer, allergic and infectious diseases

XX Example 1; Page 29; 63pp; English.

XX The invention relates to immunomodulatory oligonucleotides that comprise
 CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
 CC sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
 CC GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
 CC patients needing immune regulation, such as those suffering from cancer,
 CC an allergic disease and asthma. They are also used to prevent infectious
 CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
 CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
 CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
 CC Schistosoma. The immunomodulatory sequences are used to screen for human
 CC immunostimulatory activity by incubating macrophage cells and the
 CC oligonucleotide; and determining the relative amount of Th1-biased
 CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
 CC oligonucleotides that were tested for immunostimulatory activity. These
 CC were used in experiments for the stimulation of cytokine production and
 CC were found to lack immunostimulatory activity. The invention provides
 CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.

XX Sequence 22 BP; 5 A; 7 C; 4 G; 6 T; 0 other;

AAV80105 Length: 22 March 5, 2002 14:19 Type: N Check: 8045

Found using 'seq4' (pappu403.key)

1 TCATCTCGACGTTTCACGGTCA
 7 14

1 match found in sequence:

aaV80114 ; Oligo used in experiments for stimulation of cytokine production.
 (from "mycobacterng.seq")
 TOIG of: aav80114 check: 5550 from: 1 to: 20

ID AAV80114 standard; DNA; 20 BP.

XX AAV80114;

XX 12-MAR-1999 (first entry)

XX Oligo used in experiments for stimulation of cytokine production.

XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX Synthetic.

XX Key Location/Qualifiers
 FT modified_base 8

```
/*tag= a
/note= "5-bromocytosine"
```

WO9855495-A2.

10-DEC-1998.

05-TIN-1998:

DE TRIN. 1007.

100

NO EFFECT, TFM

PT contain an immune-stimulating octanucleotide sequence; for treating cancer, allergic and infectious diseases

Example 2: Page 30: 63pp: English.

The invention relates to immunomodulatory oligonucleotides that comprise at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS sequences are selected from the group consisting of AACGTTCC, AACGTTCC, GACGTTCC, and GACGTTCC. The immunomodulatory sequences are used to treat patients needing immune regulation, such as those suffering from cancer, an allergic disease and asthma. They are also used to prevent infectious diseases such as influenza, herpes, hepatitis B, human immunodeficiency and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and Schistosoma. The immunomodulatory sequences are used to screen for human immunostimulatory activity by incubating macrophage cells and the oligonucleotide; and determining the relative amount of Th1-biased cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent oligonucleotides that were tested for immunostimulatory activity. These were used in experiments for the stimulation of cytokine production and were found to lack immunostimulatory activity. The invention provides specific claimed examples (AAV80096-103) of immunomodulatory sequences.

Sequence 24 BP: 5 A: 8 C: 4 G: 7 T: 0 other: XX

AAV80115 Length: 24 March 5, 2002 14:19 Type: N Check: 1722
Found using 'seq4' (pappu403.key)

1
TCCATAACGTTTCGCCTAACGTTTCG
6 13 17 24

2 matches found in sequence:
 aav801116 ; Oligo used in experiments for stimulation of cytokine production.
 (from "mycobacterin.seq")
 TOIG of: aav801116 check: 1722 from:1 to: 24

TD AAV80116 standard: DNA: 24 BP

XX
AC
AAV90116.

XX
XX

XX DE oligo used in experiments for stimulation of cytokine production.

KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
 KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
 KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
 KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
 XX
 QS Synthetic.

| | Key | Location/Qualifiers |
|----|---------------|---------------------|
| XX | | 8 |
| FH | | |
| FT | modified base | |

FT /*tag= a
 FT /note= "5-bromocytosine"
 FT 19
 FT /*tag= b
 FT /note= "5-bromocytosine"
 XX
 PN WO9855495-A2.
 XX
 PD 10-DEC-1998.
 XX
 XX
 PF 05-JUN-1998; 98WO-US11578.
 XX
 PR 06-JUN-1997; 97US-0048793.
 XX
 XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
 XX
 PA Dina D, Roman M, Schwartz D;
 PI WPI; 1999-059898/05.
 XX
 DR Immunostimulatory oligonucleotides regulate the immune system - and
 XX contain an immune-stimulating octanucleotide sequence; for treating
 PT cancer, allergic and infectious diseases
 XX
 PS Example 2; Page 30; 63pp; English.

CC The invention relates to immunomodulatory oligonucleotides that comprise
 CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
 CC sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
 CC GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
 CC patients needing immune regulation, such as those suffering from cancer,
 CC an allergic disease and asthma. They are also used to prevent infectious
 CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
 CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
 CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
 CC Schistosoma. The immunomodulatory sequences are used to screen for human
 CC immunostimulatory activity by incubating macrophage cells and the
 CC oligonucleotide; and determining the relative amount of Th1-biased
 CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
 CC oligonucleotides that were tested for immunostimulatory activity. These
 CC were used in experiments for the stimulation of cytokine production and
 CC were found to lack immunostimulatory activity. The invention provides
 CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
 XX
 SQ Sequence 24 BP; 5 A; 8 C; 4 G; 7 T; 0 other;

AAV80116 Length: 24 March 5, 2002 14:19 Type: N Check: 1722
 Found using 'seq4' (pappu403.key)

1 TCCATAACGTTCCGCTAACGTTCCG
 6 13 17 24

1 match found in sequence:
 aaz11362 : Nucleotide sequence of M. vaccae antigen GVC-22B.
 (from "mycobactering.seq")
 TOIG of: aaz11362 check: 5978 from: 1 to: 554

ID AAZ11362 standard; DNA; 554 BP.
 XX AAZ11362;
 AC
 XX 25-OCT-1999 (first entry)
 DT
 XX Nucleotide sequence of M. vaccae antigen GVC-22B.
 DE
 XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KW squamous cell carcinoma; melanoma; ss.

XX Mycobacterium vaccae.
 OS
 XX WO99332634-A2.
 PN
 XX 01-JUL-1999.
 PD
 XX 23-DEC-1998; 98WO-NZ00189.
 PF
 XX 04-DEC-1998; 98US-0205426.
 PR
 XX 23-DEC-1997; 97US-0996624.
 PR
 XX 23-DEC-1997; 97US-0997080.
 PR
 XX 23-DEC-1997; 97US-0997362.
 PR
 XX 11-JUN-1998; 98US-0095855.
 PR
 XX 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
 PI WPI; 1999-430163/36.
 DR P-PSDB; AAY14903.
 XX
 XX Enhancing immune response to an antigen
 PT
 XX Claim 3; Page 204; 243pp; English.

CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
 CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.

Sequence 554 BP; 112 A; 187 C; 178 G; 77 T; 0 other;

AAZ11362 Length: 554 March 5, 2002 14:19 Type: N Check: 5978
 Found using 'seq4' (pappu403.key)

1 GATGTACGCCCGGAGAGATGTAACGTTCCGACCGGAGAACGCCGTCGCACACGAGTTAC
 22 29

61 GTTGTAGCAGCTTCAGATCT

1 match found in sequence:
 aaz41866 : IL-12 secretion inducing CpG oligonucleotide 11.
 (from "mycobactering.seq")
 TOIG of: aaz41866 check: 5318 from: 1 to: 20

ID AAZ41866 standard; DNA; 20 BP.
 XX AAZ41866;
 AC
 XX 24-JAN-2000 (first entry)
 DT
 XX IL-12 secretion inducing CpG oligonucleotide 11.
 DE
 XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;

KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 OS Synthetic.
 XX
 PN WO9951259-A2.
 XX
 PD 14-OCT-1999.
 XX
 XX
 PF 02-APR-1999; 99WO-US07335.
 XX
 PR 03-APR-1998; 98US-0080729.
 XX
 XX (IOWA) UNIV IOWA RES FOUND.
 PA
 XX
 PI Krieg AM, Weiner G;
 XX
 DR WPI; 1999-620169/53.
 XX
 XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system
 PT
 XX
 PS Example 8; Page 69; 91pp; English.
 XX
 CC Sequences AAZ41856-Z41949 are phosphorothioate Cpg oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX
 SQ Sequence 20 BP; 4 A; 7 C; 3 G; 6 T; 0 other;

AAZ41866 Length: 20 March 5, 2002 14:19 Type: N Check: 5318 ..
 Found using 'seq4' (pappu403.key)

1 ATCGACTCTCGAACGTCTC
 |-----|
 10 17

-- Search Statistics --

Times: CPU Total Elapsed
 00:00:01.02 00:00:02.00

Number of sequences searched: 880
 Number of sequence hits: 13
 Number of separate matches: 16
 Number of sequence hits saved: 0


```
...
6398 GCAGGAACGGGAAGATTTACCATAACTCCGGTTATCAGTACCATCGGCTCAACGCTCGTT
      |-----|
      6448 6453
6458 GTCCGATCTGAATAATTCGCTCAAAAGATCATATTTCCCTGGATATTTTCCACCGTTCT
      |-----|
6518 TATGTGAGCAAACTCACATAATTTCTGTACAGCAGCAGAGAAAACGGATATCGATTATTTGTT
      6547 6552
6578 TAATATTTTACATTATTAATAATG
...
6822 ACGCAGAAACAGGAAGAAAGATCAGCCAGTTAGACTGGAAGATCAAAAATGTGCGCTATCC
      |-----|
      6872 6877
6882 TGAAGGTGATATATCCTGGGATCCATCTACTCATTTTCTGACCCTGAATGCCAGGGGTGGA
      |-----|
      6940
6942 CGTCTCTGGCTTCGGGTCAGGTAATATGATGACTACGACTGAGTGAATGAAA
      ---|
      6945
...
7183 AGCTTATACCGGAAACTTCCCGAAAGAGTGGGGTAATAGGTTATAACCAAGCGCTTTC
      |-----|
      7233 7238
7243 TATGCCATATNTGGACTTGCAGGCCAGTATGCCATTAATGATTTTGAGTTAAATGCATT
...
7303 ATTTAAATTCACGCACTGGGTTCCGGCACATGATAATGATGAGCACTATATGAGAGATCT
      |-----|
      7313 7318
7363 TACTTTCCGTGAGAGACATCCGGCTCACGTTATATGTACCGTAATTAACGCTGGATA
      |-----|
      7412 7417
7423 TTATGTCACACCTAATGCCAAAGTCTTTGCGGAATTTACATACAG
...
7543 TGCTGCCGGTATTTCCAATAAAAAATTAATGTGACGGCGGGTCTGCAATATCGCTTCTG
      |-----|
      7593 7598
7603 AAAAAACAGATCATATCTCTTTTTCATCCTCCCTAGCGGGGAG
...
7676 TGACCAACCTTCAGATGTGTGAAAAATCACCTTTTTCACCATAAATGACGGGGCGCTCAATT
      |-----|
      7736 7731
7736 CTGTTGTTTGGCTTGACATTTCTCCACGTCTTTCAGGGCATGGAGAAGGTCAAAATTAGAC
...
7796 ATGGAACGCTACTCTCTCTCTGTAGGAAGCTCAACATCCAAGCTTAATTTGCTCCCAT
      |-----|
      7800 7805
...
7856 TGCCTTCAAGCTTAACGCTTAACGCTCGGCAGCTTTAAATCATTTCCGGCTGCTCCAGCTT
      |-----| |-----|
      7867 7872 7878 7883
```

```
7916 TGTACTGCTGGCTTATACCCATCGCTCAGCAACTTGTJTTTTGTGATAACTGGAG
      |-----|
      7941 7946
7976 TTCTTCAGGCATCATCTGCAA
...
8006 AAGAATCATCTCATCTGCCAATTTCTTTAATTCGTGTCTGGCTTTCAGGTGAACGACTGGC
      |-----|
      8056 8061
8066 AATCACTCATCTAATGTTCTCTCATTTACTTGTCTCTCCAGTGTGTCA
...
8144 ACCAGTTTTTCATAAAACCGCTTATCATTTACTTTTATCTCTGCACAAAAGCATAGCC
      |-----|
      8194 8199
8204 CGACGAATCGGATCGAAGCATAAAGGCTCTTATCGGACGGCCAG
...
8327 GGTCGCTAAACCTGTAGCTTTTTCAAATCAGCCAAAACCTTTTCTCTGAAGAGCGCTTCT
      |-----|
      8377 8382
8387 TGCTCATTTAGCCAGTCGTCAAAATCGCTGGCTAAAAAGTACCATCCACATGCTCAACCT
      |-----| |-----|
      8401 8406 8414
8447 ATAACTGTAGCTTACCC
...
8942 GCCACAGCTCTGTCTGTGTGGCAGCTTTTCGCTGTGCCGAAACCTGCTTAAACGCTCCCG
      |-----|
      8992 8997
9002 AAAGGCCCTCTGAACCAAGAACAAACACACAGGCCATTAAAGTAA
...
9075 GCTGCAAAAAAAGTCCCTAATGGAGCAGGGACTGTTAAACCCAGTGAATAGGCTCTAAA
      |-----|
      9125 9130
9135 TTAAAGTAGAATACGACCAGGTACTCTTTCAGAAAAAGAGATTATC
...
-- Search Statistics --
Times:      CPU      Total Elapsed
           00:02:54.10      00:02:57.00
Number of sequences searched:      90
Number of sequence hits:          85
Number of separate matches:      7870
Number of sequence hits saved:    0
```



```
/transl_table=11
/label=yppCp1.09c
/product="hypothetical protein"
/protein_id="CAB53172.1"
/db_xref="GI:5763819"
/translation="MMVLFQSOFDDMLNEQEDALQEKVLADLKKLQVYGPPELPRPYA
DTVKGSRYKNMKELRVQFSGRPIRAFYAFDPPIRAIVLCAGDKSNDKRFYEKLVRIAE
DEFTAHNLTLSEK"
8528..8532
/misc_feature
  /note="difference: AF053945 has 3 T's this sequence has 5"
  /label=diff
BASE COUNT      2792 a      2253 c      2098 g      2469 t
ORIGIN
YppCp1 Length: 9612 February 21, 2002 16:59 Type: N Check: 5414
Found using 'seq2-3' (pappu403.key)

1  TGTAAAGAAAGGTGCAATAGTAGTCCACACCAACCGCCTGAAATCAGATCCAGGGGGTAA
33 38
61  TCTGCTCTCCTGATTCAGGAGAGTTAT
...
99  GAGACAGTTATGGAATTAATAATCCTGTCACAGCAGGGGAATGAGTAGCCGGCGGATGCC
149 154
159  AGAGAACTGGGATCTCCGGCAATACCGTTAAACGTTATTTTGCAGGCAAAATCTGAGCGG
190 195
219  CCAAAATATACCGCGGACCTGCTGTTGCTTCACTCTCGGTGGAATGAATACCGGGATTATAT
288 293
279  CGTCAACGATCGCGGATGCTCATCTTACAAATCCCGGCAACGATATCTGCTCGCGAG
327 332
339  ATCAGACACAGGGATATCGTGGCGGAATGACCATTCACAGGC
...
398  CTCGGTTCTCTCAGGACGAGGACCTCGCGTTTCGGTTTCGAAACTGAACCGCGGACGACAGAT
448 453
458  GCAGGTTGACTGGGGCACTATGCGTAATGGTCGCTCACCGCTTCACGTTGCTGCTGT
487 492
518  TCTCGGATACAGCCGAATGCTGTAC
...
701  TCCTTCGCTGTGGCAGTTTCGGCAAGGAGATGGGCTTCTCTCCCGCAGCTGTGTCGCCCTTT
751 756
761  CAGGCACAGACTAAAGGTAAGTGGAAACGGATGTTGCAGTACACC
...
859  CTGTCGATGTTGAAACAGCCAAACCGCACGGTCTCGGCTGGCTGCACGATGTCGTAACC
909 914
919  AACGAAAGCATGAACAATCCAGGCCCGTCCCTTGGGATCGCTGGCTCGAAGAGCAGCAGT
955 960
979  CCATGCTGGCACTGCTCCGGAGAAAAAAGAG
```

```
...
1078  TCTACGACTCATCTCCAGAGAGTGGCGTGATGGAACATCCACATCAACGACTGAT
1128 1133
1138  GGCGTTCGCGCGGCGAGTTGCAACTGGAAAGCCTTATAAGCGCAGCGCTGCGCTGTCA
1138 1143
1198  ACAGGAGTAGACCAGGAATGGAGTTATATGGACTTCCTGGAGCATCTCGTTCATCAAGA
1258  AAAACTGGCAGCTCATCAACGTAACAGCGGATGTATACCCGAATGGCAGCCTTCCTCCGGC
1285 1290
1318  GGTGAAAACGTTTCGAAGAGTATGACTTTCATATCGCCACCGGAGCACCAGCAAGCAACT
1324 1329
1378  CCAGTCGTTTACCGCTCACTCAGCTTCATAGAACGTAATGAAATATCGTATTACTGGGGCC
1381 1386
1438  ATCAGGTGTGGGAAACCCATCTGGCAATAGCGATGGGCTATGAAGCAGTCCGTCGAGG
1468 1473
1498  TATCAAAGTTCGCTTCACAACAGCAGCAGATCTGTACTTCACTTATCTACGGCACACG
1554
1558  TCAGGCGCGTTATAAAGCAGCGCTTCAGCGTGGAGTAATGGCCCCCGCGCTGCTCATCAT
1559 1578
1576 1581
1618  TGATGAATAGGCTATCTGCCGTTTCAGTCAGGAAGAAGCAAGCTGTTCTTCAGGTCAAT
1676
1678  CGCTAAACGTTTACGAAAGAGCGCAATGATCTCTGACATCCAATCTGCCGTTCCGGCAGTG
1681 1688
1738  GGATCAAACGTTTCGCCGTTGATGCAGCACTGACCTCAGCGATGCTGGACCGTATCTTACA
1744 1749
1774 1779
1798  CCACTCACATGTCGTTCAAAATCAAAGGAGAAAGCTATCGACTCAGACAGAAAGAAAGC
1808 1813
1858  CGGGGT
...
1895  GGATCAATATTGGGCGGTTGGTGGAGATATAAGTGGATCACTTTTCATCCGCTGTTGACA
1945 1950
1955  CCCTGATGAATTCACGTTTCACGCTGAATAACAAGATGCCGA
...
2002  ATACGCACTCATATTTTTTTTACACAATCTCTTAATCCCGACAAGTCTGTATATAG
2052 2057
2062  GAAAATCTTAGCACCATTCGCGAACAATCAGAACAGCAGGCCATGAACGACTGACAACA
2108 2113
```

| | | | | | | | | | | | | | | | | | |
|--------------|---|-----|---|------|--------------|--------------|------|-----|------|-----|------|-----|------|-----|--------------|--------------|------|
| misc_feature | 1433..1456 /gene="yppCp1.02" /note="PS00017 ATP/GTP-binding site motif A (P-loop)" 2285..2283 /label=rep 2321..2340 /label=rep complement(2413..2432) /label=rep 2924..4176 /note="similar to ColE1 ori" /label=ori 2925..3119 /gene="yppCp1.03" /note="rop" 2925..3119 /gene="yppCp1.03" /note="yppCp1.03, rop, plasmid replication regulatory protein, len: 64 aa; highly similar to e.g. ROP_ECOLI regulatory protein ROP (RNA one modulator) from ColE1 (63 aa), fasta scores; opt: 241 z-score: 482.0 E(): 1.5e-19, 63.9% identity in 61 aa overlap." /codon_start=1 /transl_table=11 /label=rop /product="putative replication regulatory protein" /protein_id="CAB53166.1" /db_xref="GI:5763813" /translation="MNKQQTALNMARFTRQSLLILEKLDALDADEQAAMCERLHEL AEPLQNSIQARFEASETGT" 4355..4780 /gene="yppCp1.04" /note="pim" 4355..4780 /gene="yppCp1.04" /note="yppCp1.04, pim, pesticin immunity protein, len: 141 aa; identical to TR:Q57285 (EMBL:Z54145, X92856)" /codon_start=1 /transl_table=11 /label=pim /product="pesticin immunity protein" /protein_id="CAB53167.1" /db_xref="GI:5763814" /translation="MISKLFCLALIFLSSGLAEKNTYAKDILONLENTFGNSLSH GIYQKTFKQTEFTNKSNTKKHIALINKDSWMSLKILGKRDEYVCFDEFSLI RPTVAIHPLLIKVKSGNFIVVKEIKKSIPGCTVIYH" complement(4815..5888) /gene="yppCp1.05c" /note="pst" complement(4815..5888) /gene="yppCp1.05c" /note="yppCp1.05c, pst, pesticin, len: 357 aa; identical to TR:Q57159 (EMBL:X92856)" /codon_start=1 /transl_table=11 /label=pst /product="pesticin" /protein_id="CAB53168.1" /db_xref="GI:5763815" /translation="MSDTMVNGSGVPALFSGSTLSSYRPNFEANSITIALPHYVD LPRGNFKLMYINGFPIDTEMKDESYSNKIRQESKISTEGTVSYEQKITVETGOEK DGKVVYRVMLVGETTAESIEHLDKKENEDILANNRRIRIVLADNTVINFONIQLKEFL RRSVNLVDHIDIFSSNGFEGFNPTSHFSPNPSDDYNSTGVTGSGVDLQQRKQDLLN DGVPQYIADRLDGYMLRKEAYDKYRTPAPLTSNEAHLNIIYIDRFSHKIEGLFN DANIGLRFSDLPRTALTALVSGYQKGFKLSRTAPTWNKVIATKDMNGLVNAFNIVD GMSDRKRREGALVQKIDISGLLK" 5910..5917 /note="difference: AF053945 has 7 A's this sequence has 8" /label=diff 5927..5934 /note="difference: AF053945 has 9 A's this sequence has 8" /label=diff 6006..6422 | CDS | /gene="yppCp1.06" 6006..6422 /gene="yppCp1.06" /note="yppCp1.06, unknown, len: 138 aa" /codon_start=1 /transl_table=11 /label=yppCp1.06 /product="hypothetical protein" /protein_id="CAB53169.1" /db_xref="GI:5763816" /translation="MKPHFCDLNHSYKNOEKGIBSRKTAPCNIBKKOKGDMVSKTKSG RHLSKTDKRLAALVAVGYEERTARDLIQKHVYTLTQADLRHLVSELSNGVGOSQAY DAYQARRIRLARKYLSGKAPEGVPEGREGREDLP" 6665..7603 /gene="yppCp1.07" /note="pla" 6665..7603 /gene="yppCp1.07" /note="yppCp1.07, pla, coagulase/fibrinolysin precursor (EC 3.4.21.-) (plasminogen activator), len: 312 aa; identical to COLY_YERPE. Contains PS00834 and PS00835 Serine proteases, ompT family signatures 1 and 2" /codon_start=1 /transl_table=11 /label=pla /product="coagulase/fibrinolysin precursor" /protein_id="CAB53170.1" /db_xref="GI:5763817" /translation="MKKSSIVATITILSGSANAASSOLIPNISPDSTVAASWTGMLS GKSHMELYDAETGRKISOLDWKIKNVALKGDISWDYPSEFLTINARGWTSLASGNGM DDYDMNENQSEWTDHSSHPATNHNANEXDLNVKGLLQDENTKAGITAGYQETRES WTATGGSYNNAGATYGNFPKGVGVIGYQNRFSMPYIGLAGOYRINDFELNALFKFSD WYRAHNDHEHYMRDLTFREKTSGRYGYTVINAGYVYTPNAKVFAPAEFTYSKYDEGKGG TQTIDKNSGDSVSIIGGDAAGISNKNYTVTAGLYRF" 7007..7036 /gene="yppCp1.07" /note="PS00834 Serine proteases, ompT family signature 1" 7118..7168 /gene="yppCp1.07" /note="PS00835 Serine proteases, ompT family signature 2" complement(7790..8089) /gene="yppCp1.08c" complement(7790..8089) /gene="yppCp1.08c" /note="yppCp1.08c, probable transcriptional regulator, len: 99 aa; similar to several predicted transcriptional regulators e.g. TR:P95258 (EMBL:Z84498) Mycobacterium tuberculosis MTCY09F9.0 8C (RV1956) (149 aa), fasta scores; opt: 190 z-score: 314.7 E(): 3.1e-10, 35.4% identity in 99 aa overlap. Contains helix-turn-helix motif from aa 42-63 (Score 2083, +6.28 SD)" /codon_start=1 /transl_table=11 /label=yppCp1.08c /product="putative transcriptional regulator" /protein_id="CAB53171.1" /db_xref="GI:5763818" /translation="MRTLDEVIASRSPQTRIKEMADEMILEVQLQMMREELQLSQK QVAEAMGISQPAVTKLEQRGNLDKLATLRKYVEAMGGKSLDVELPTGRRVAFHV" complement(8089..8436) /gene="yppCp1.09c" complement(8089..8436) /gene="yppCp1.09c" /note="yppCp1.09c, unknown, len: 115 aa; similar to two hypothetical proteins from M. tuberculosis TR:O53468 (EMBL:O53468) MPTV018.09C (RV2022c) (201 aa), fasta scores; opt: 271 z-score: 398.5 E(): 6.7e-15, 41.7% identity in 108 aa overlap, and TR:O53332 (EMBL:AL021646) MTV014.26 (RV3182) (114 aa), fasta scores; opt: 214 z-score: 320.5 E(): 1.5e-10, 39.4% identity in 104 aa overlap" /codon_start=1 | gene | misc_feature | misc_feature | gene | CDS | gene | CDS | gene | CDS | gene | CDS | misc_feature | misc_feature | gene |
|--------------|---|-----|---|------|--------------|--------------|------|-----|------|-----|------|-----|------|-----|--------------|--------------|------|

```
...
331211 AAATTTTACTCTTCTTGGGAACATTATATCGTACCGGCATCGATTCCCAAGCGCCCAT
|-----|
331261

331271 TCCGGCGGTGCGGTTTAAACGGCGGTTTATGTATATATCGCGGGTTGTGGGCGTATCC
|-----|
331308

331331 GATGGACTTGGCGATTGGCAGCGGATTATGACGGTAGCCTTGGCTGAGGCGTATACTT
|-----|
331340

331391 ACCGC

...

331446 GCCGCCACCGTGTGTTCTCTCCGCTTGGTCAATCCGTTTTCGGCTGGCGGTTCACG
|-----|
331496
331502

331506 ATGCTGTGATAATTGGCTTAATCGGCTGCAAGAGCGCAGCGGCAAT
|-----|
331507

...

331563 TTCACCGGCGCGTATTGATACCGTAGTAGTTTCTTGATCTGTGTGGCGATGGGT
|-----|
331613

331623 GCGGTGGGATGCTGCCCGGTATCCGCGGTTTTTGGAGCACTCA

...

-----
78 matches found in sequence:
yppcp1 : TOIG of: yppcp1 check: 5414 from: 1 to: 9612
(from "mycobacter.ge.seq")
TOIG of: yppcp1 check: 5414 from: 1 to: 9612

LOCUS      YPPCP1      9612 bp      DNA      BCT      19-AUG-1999
DEFINITION Versinia pestis plasmid pPCP1.
ACCESSION  AL109969
VERSION    AL109969.1 GI:5763810
KEYWORDS   coagulase; fibrinolysin; ISI00; omptin; pesticin; pesticin
            immunity; p1m; pla; pst; rop; transcriptional regulator.
SOURCE     Versinia pestis.
ORGANISM   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Versinia.
REFERENCE  1 (bases 1 to 9612)
AUTHORS   Karlyshev,A.V. and Wren,B.W.
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 9612)
AUTHORS   Baker,S.G. and Mungall,K.
JOURNAL   Unpublished
REFERENCE  3 (bases 1 to 9612)
AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE     Direct Submission
JOURNAL   Submitted (19-AUG-1999) Versinia pestis sequencing project, Sanger
            Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA
E-mail:   barrellesanger.ac.uk DNA supplied by Dr. Andrew Karlyshev
            and Prof. Brendan Wren, [3]. Department of Infectious and Tropical
            Diseases, London School of Hygiene and Tropical Medicine, Keppel
            Street, London WC1E 7HT
Notes:
Versinia pestis sequencing at The Sanger Centre is funded by
```

```
Beowulf Genomics
Details of Y. pestis sequencing at the Sanger Centre are available
on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/Y_pestis/)
CDS are numbered using the following system eg pPCP1.01c. YP (Y.
pestis), PCP1 (plasmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
CAUTION: We may not have predicted the correct initiation codon.
Where possible we choose an initiation codon (atg, gtg, ttg or
(att)) which is preceded by an upstream ribosome binding site
sequence (optimally 5-13bp before the initiation codon). If this
cannot be identified we choose the most upstream initiation codon.
FEATURES
            Location/Qualifiers
source      1..9612
            /organism="Yersinia pestis"
            /plasmid="pPCP1"
            /strain="CO-92 Biovar Orientalis"
            /db_xref="taxon:632"
repeat_unit 1..1954
            /note="ISI00"
            /label="ISI00"
gene        87..1109
            /gene="YPPCP1.01"
            87..1109
            /gene="YPPCP1.01"
            /note="YPPCP1.01, probable ISI00 transposase, len: 340 aa;
            identical to TR:P4993, and similar to e.g. TRA6_BACST
            putative transposase for insertion sequence element IS5376
            (400 aa), fasta scores; opt: 632 z-score: 1097.4 E(): 0.
            38.1% identity in 328 aa overlap. Contains
            helix-turn-helix motif from aa 19-40 (Score 2045, +6.15
            SD)"
            /codon_start=1
            /transl_table=11
            /label="YPPCP1.01"
            /product="putative transposase"
            /protein_id="CAB53164.1"
            /db_xref="GI:5763811"
            /translation="MVFETVMEIKLHKQMSRAIARELGISRNTVRYLOAKSEP
            PKYTPRAVASLLDEYRDIQRIDADAHPIKIPATVIAREIQDQYRGOMTILRAFIR
            SLSVPEQEPAPVAFETEPGRQMDWGTMRNGRSPLVHFAVLGSRMLYIEFTDMNR
            YDTLETCHRNAFRFFGGVPREVLVDNMKTIVLQRDAYTGQHRFHPSLQFGKMGFS
            PRLCRPFRAQTGKVERMVQYTNSFYIPLMTRLRPMGITVDVETANRHGLRLHDVA
            NORKHETIQARPCDRKLEEQQSMALLPPEKKEYDVHLDENLVNFKHPLHPLSIYDS
            FCRGVA"
            1106..1888
            /gene="YPPCP1.02"
            1106..1888
            /gene="YPPCP1.02"
            /note="YPPCP1.02, ISI00 ATP-binding protein, len: 260 aa;
            identical to TR:P4994 and similar to e.g. ISPB_ECOLI
            insertion sequence IS21 putative ATP-binding protein (265
            aa), fasta scores; opt: 730 z-score: 1078.9 E(): 0, 47.4%
            identity in 249 aa overlap. Contains PS00017
            ATP/GTP-binding site motif A (P-loop)"
            /codon_start=1
            /transl_table=11
            /label="YPPCP1.02"
            /product="putative ATP-binding protein"
            /protein_id="CAB53165.1"
            /db_xref="GI:5763812"
            /translation="MMMLKHORLMALAGQLQLLESLSAAPALSOAVDQENSYMDFL
            EHLHHEKILARHQRKOAMTYRAAFPAVKTFEYDFTFATGAPKQLQSLRSLSFTLR
            ENIVLLGSPGSKTHLAITAMGYEAVRAGIKVRFYTTAADLLQLSTAQRGSKYTLQ
            RGVMAPRLLIIDEIGYLPFSQEEAKLFQVIARKEKSAMILTSNLPFGQWDQTFAGD
            AALTSMALDRILHSHVVQIKGESYRLRQKRKAGVIAEANPE"
```


325080 GTACTGCTCGATCGGCTGCCAATCAATCACTGGTCCAACTTCAAT 325070
...
325311 TTTAAGTTTGACCGGTTTCAATCATCGGTATTATGAAATTTCAACATCGCCAAATC 325361
325371 TATCCITTAATCTCTTTTCCAAATTCGCGAGATTCGGCGGAAAGTTT
...
325659 CCAAGCCCGATGCGTCGATTAAATATAGTGATTAAACAAAATCAGGACAAGCGACGAAG 325709
325719 CCGCAGACAGTACAAATAGTACGAACCGATTCACTTGGTGTTCAGCACCTTTAGAGAAT 325777
...
325779 CGTCTCTTTGAGCTAAGCGGAGGCAACGCCGTACTGTTTGTGTTAATCCACTATACCA 325804
325839 CCTTGTCTTTATTGGCTTTAT 325843
...
325892 TAGCCCTTTTGGTGTACCGGCTTGGCATACTGTTCTGCTTCAGACGGCATCGCTTCAT 325942
325952 TTGCGCTTAAATACTTCTTCGTCCAGCGATTTCAAACCACGCCAGCTTTTCGCCGATTTG 325976
326012 ATTTCCAGTCCGCGGGAC 326027
...
326042 AAATCGGGTTCGTCCAGCGGCTCGGCATATAGCTTTCGCCGGCGGTAGCGGTTTCGGT 326092
326102 TCGTGTGGCGTAGCGGTATTCGCGTCCGTAGCCCAATTCCTTCATCAGCTTGGTCGGG 326161
...
326162 GCGTGGCGAGGTGGACGGCACTTCGTCGCTGGCGTTTCTTTGACGAAGTGGCGCAT 326188
326222 TGGTGTATGCCTTGTAGCCCGGTTGG 326194
...
326266 AAGATACAATACCGCTTGCCCAAGCCAGTTTCGCCCTTCGGCGAGCCTTAAGCGTTCGAA 326316
326326 GGTGGCGGCGCATCGTTGGCGATTTCGAAGCGCGGGTCGGCAAGCCCGATGCTTC 326338
... 326344
326386 CCAAGCGATACGCACGATGCGGGCGGAGGTAGCGGGTTCGGTCCCGCGCTCGAGCAT 326389
...
326446 ACGCAGAACCAATACAGCGCGGCTTCGGATGCGAACCCGCGACGGATTTGTCCAGGGC 326467
326506 GGAGATTTGGTTGTAGA
...
326551 GATTTGGCCCCGAGACTGTTCGGGGAGAAATTCGGCGGTTAAGTTTTCAGACACAGTGT 326601
326611 ATCGGCGCGGTAAAGTTGTTTCCAACAAATTCACAATCTGCGC
...
326711 AACTCTTGGTATTTCAGGCAAGCAATACCTTGGCAATCAGCTTTTTCAGGTCGCTGAA 326761
326771 GACAAGGTTTGCAAAACATACACCTGAGCGCGGCTCAACAGCGGGGATTTGACTTCAAAC 326828
326831 GACGGATTTTCCGCTCGTCGCCCAATAAAGTTAGCAAAACCGCTTTCGACATCGGCAAA 326833
326891 AAGCGCTCTCTGCTGCGCCTTGTGAAGCGGTGGAATTCATCGACAAAACAAATCCCGGCC 326942
326951 CGTCCCTGCTGCAAGCGATTTTCGGGTTTATGATTCGCTTCGGGATGTCCTTCACGCGC 326965
327011 GAAATACGGCGGAACAGCAAAACTGGGCGTTGAAACTCTGCGCCAAATCCCGGCC 327040
327071 AAGTCTGCTTTCGCCACGCCCGCGGCCCAACATAGAAATGCGGCTTTCGCCGC 327071
327074
...
327137 GCCACGGCAAGGTTTACCTTCGCCGATGAGGTGTTCTTCGCCACCACGTCGTCAAGC 327187
327197 GTATCGGACGCAATCGTTTCGGCAAGCGCGCGTAGGGTTTCGGGCAAAACAAATCGGTC 327210
327257 ATAACGGCT 327210
...
327397 TATATCCGGCAGGACCGGATTCCGACCGGATGACAATACCCAAATCAGGCTTCAA 327447
327457 TTAACATTAAAGGAGACTAAAATAGAAAATTTGCTTTTATCTACCAT
...
327556 CTAATATCAAGAGATGGAATGGGTGCTTTATTTTCTGATCGGCAAAAGAGCATGATA 327606

```
321271 TTTACAATGTGAAAAATACCTCCGCTGCTCGGTGGATTCCGTGCT
...
321412 CGCATATAAAGTGATTTCATCAAGAAACGGCGGCTGTCGGATGCCGGAACGCCGGTA
|-----|
321462

321472 TCCGGCGGCAAAAGCGGATTACCTAATCTTTTGGACAGGACGACTATTGGGCGGATA
|-----|
321486
|-----|
321510
321513

321532 CCAACCGTTCAAAAAACGCGGGGGGGGATTCTCTTT

...

321638 ATCCCAAAGGGCGGACTTTTCGGATAATGATTTTGTCCGCCATTTTGAACGTTGGTG
|-----|
321688

321698 GAGGGCGGTACTATATCGCCCAACGCGTGGACAAAGATTGTACGCGGGAAATATCATC
|-----|
321713

321758 AAAAACAATCT

...

321870 ACCGGTCTCGGCAACTCTATACGCCACACATCAATACAAAAAATTCACGGATATGC
|-----|
321920

321930 TGACGCATCTCGACCGGGCGTGGATTTTTAGTCGAAACAAAAA

...

322020 TCGGTATCTGAGATCTATATCGGTAAGGCTTATTTTCCAAAAACATTATCGTTGTCT
|-----|
322070

322080 ACCGAGAATATTTTCATTTAAAGAGAAATGCAGAAAGATATTCCG

...

322419 TGGTAAATACCGGATTACAATCCGCTACATCCGATTCTACAAGGATGAACGATGACC
|-----|
322469

322479 GACACGCCGTCTGCGCCCGCAACCTGCGGCAATGGATAGCAG

...

322762 AAGTCGGCAAAATGGTCATCATACCGATGCCGAAGACCGAGAAACGAGCGACCTGC
|-----|
322812

322822 TGATGGCGGCGCAATTCTGTCACGCCGGAAGCCATCAACTTCATGAT

...

323032 GCGCCTGACTATTCAAACCGCGCTTCCCGACCGCCCAACCCGGAAGACATCGTCCAAC
|-----|
323082

323092 CCGGTATATCTTCCCGTTCCGCCCAAAAGGGGGGTACTCGTCGCGCGCGGACACA

323152 CCGAAGCGCGGTGCGACTGCGCAAAATGAACGGGCTGATTCTCTGCCCGCTTATTTCGG
|-----|
323160
|-----|
```

```
323212 AAATCATCAACGACGAGCGCACGATGGCGGTATGCCCGAACTGATGAAATTCGCCGAAG
|-----|
323220
323223

323272 AACACAA

...

323527 ATTCTGTGTCATTGCCCAAGCACTCGAGCGGTCCACACAGCAGAAAGCGGCTCGTCA
|-----|
323577
323580

323587 TCCTTCTGCACCGCACCGAAGACGGCGCATCCCTGCTCGACCGCACCCCT

...

323662 AATGGGACAGCAAAAGCTACGGCATAGCGCGCAAAATCTCTCGCGGCTCAACGTCAAAA
|-----|
323712

323722 AACTGCGGTCCTCGGGCAGCCCTCATCTTTTACCGGCTGACCGGCTTCGGTTGGGAG
|-----|
323781

323782 TCGTCTGGCTTTGAAGAAGCGGAAAAATAATAGTAAATTCAAATACTTTATATT
|-----|
323786

...

324189 CGGCAGCGGGCGCACTTAGCCTGTTGGCCGCTTTCAACAGGTTCAACACATCGCTTCA
|-----|
324239

324249 GGTGGCTTTGCGCACTCACTTTAATCATGTCGGAATAAGGCTGCCCG

...

324477 AGCCTTTGTGGGCATAGACGGTCGTACCTTCGGGTAAACCTTCCAACAACGCGACAGGT
|-----|
324527

324537 GTTTGCACCTCATGGGCATTGGCGGAGTAATGTCAGTTTTCGATATCGCTTCTGCTG
|-----|
324583

324597 CGGTACGGGTATGTTGTTGTAAACCGAGTTTGTAGAGGCCGT

...

324673 CGGTGTGGTTGGCTGCTGACTTGTCTCTCTTCGTCAACTTCTATGGCCTGGCGCTGTTT
|-----|
324723

324733 GCTCCGCGGCTCTGAATAATGGTGGGTCAACGACGCGCGGATGCTTCTCTACTTT
|-----|
324757
324763

324793 TAGCCCTTTTTCGGTCAGTTGGCAGTTAATCAGTTTGAGCAATTCGGACAGGGTCTGCTC
|-----|
324847

324853 TTGCGCCCGCGTTTCGGTAGCGGCATAAGGTGCTGTAATCGGGGATGCT

...

325020 CAGCGGTAGCGGGACGGCGCGGTGGTCTCGGAGGTAACGGGTCTTTGACGATTCAG
```

317230 TTCTGCTGAAAGCATACGCAAGCGCTGAAGCGGATAATGATTCTCGCCCGAGATGA
|-----|
317262

317290 CCGCCAAACTCGTCAAAAGCCTGATTTC

...

317365 GTGAACCTGAAATCTTGGGTATCTCGCCGCGAGGACAGCAACAATAATCATCGCCGCC
|-----|
317415

317425 ACCTCGATCTTGCCGAATCCACCGTCAAAAGTCCAGTTCAAAACCT

...

317479 AACTCAACCTCAGCAGCGGGTGCAGCGCGCGTTTACGCCATCGGCAACAGTCCGCC
|-----|
317529

317539 AACCTGTCGGGAATAGCGGTTTCAGCGGCATATTAGGGGTTTAAATCCCGTAGGGTCA
|-----|
317555

317599 TTCGGATAACAG

...

317780 TCGGGGTAAATGTTCATTCGAATCCCATAAATATACAAATGGCTTATATCGTTTAAAG
|-----|
317830

317840 CATGTGTAAACCACCCTCATATCAATATATAGTGGATTAAACAAAACCACTACGGC
|---|
317897

317900 GTTGCTCGCCTTGCCGTACTATCTGTACTGTCTGCGGCTTCGTGCGCTTGTCTCTGATTT
|-----|
317902

317960 TTGTTAATCCACTATAAAATGCCGTCCAAACCCATGT

...

318008 ATTTGCTATAGAAGCAATCAGCAACCTGGGTTTGATGCTGCTCTCCCTGACGCTCAGG
|-----|
318058

318068 GATCAAACTAAAGGTAACCTGTTTCACTTTGTTTACCCCAAGAGACCCCTGAACCCGATC

318128 GGCATCTTCGCAGCAACAATAACGACCATGCCGATGCCGAGTTAAAGGTTTCGGTACAT
|-----|
318149

318188 TTCTTGGGTTTCCACAT

...

318480 AGGGAATAGCGTTTGAATGTGCGCGTTGGAAGCCAAACCCAAATACCACATCGCCCGCG
|-----|
318530

318540 CCGATGCTGCGCGGTAATGACATTTCTCTTTTTTCCACTACGCCGA

...

318735 CAGCGGAATAATCCAGAAAAACAGAGTTCTGCTCTTGAACCAAAAATATCGTTGACA
|-----|
318785

318795 CTCATTGCAACAAGGTCGATGCCCAACCGTATPCGTGTTTATCCCAAT

...

318957 TCCGACGCATGGTCCGTTTGGCAACGGTTTGATGTTTTCGACCAGTTGTCGCGCTCGG
|-----|
319007

319017 TCGATATCGACACCTGCATCGGGTAACCTCAATGAAGTACTCATCGTTTTTCTTGGTAA
|-----|
319059

319077 ATGGGGATTGGACGGTAAATAACGGGGCGTATTCTAC

...

319462 CTGCGTGTGAACACATATAGGCAGATAAAAAAGCGCCGCTGAAAACGACAGCACTTAT
|-----|
319512

319522 GTTTTGTGGCACTAATTTGTCCGATAAGCATTAACATATATAATTTATTATCATATTG

319582 GTCGGACGGAGAGACTCGAACTCTCACACCTCTCGGCGCCAGAACCTAAATCTGGTGCG
|-----|
319617

319642 TCTACCAATTTGCCACGTCGCCATGGGATTTGACGATTTATACAGATTTTGTTTTTTGG
|-----|
319675

319702 TGCAAGGTTTTCGGCGGGGCTGTTGATGG

...

319889 GCGTATCTGACGGCGCAGACGGTAATGGATATGGGAGCGCTTCGATTGACGTTCCGT
|-----|
319939

319949 TTTGTGGCTTGAGCATGAGGATTGGAGGCTTTCGGGATTATTTGGAG

...

320456 GGGCATCACTCCCGCAGAGAGATGAAATGTCTGAATTTATAAACTATCTCGCTTCGT
|-----|
320506

320516 GCCGCTGCGCAGCGAATGGGGCGTGTACGATGCACGGTTTGAAGAAGCAGCGGGC

320576 AGSAACACGTGCGCTGACCGTCCGGGATTTTCAGACGGCAATCCGGTTCTGACGCGCA
|-----|
320599

320636 TCCACTCGGATGCTGACGGCGCAGCGGCTGTCTCAAAAAATGCGACTGCGGGCGCG
|-----|
320656

320696 AACTGGAAGCGGCAATGAAAGCGTTTCAGGCGAAGGCGGCGCATCATCTCTATCTGC
|-----|
320743

320756 GTCAGGAAGGACGCGGCATCGGGCTGATTACAAAAATCCGCGC

...

320976 TAACGTGTCGAACGCATCCCCCTGCAGCTCGGGGAAAAACCTTGAAAAACAACCGTACCT
|-----|
321026

321036 CCAAAACCAAGCAGACAAGCTGGGGCATCTGATGTCGGAATAAGGC

...

321211 ACGCGCAACATCTTAATATAATAGGATTGATTTTATGATTTTCTCCATCATCTCCCTA
|-----|
321261

```
313604 TGTTGATCTGTCGCCGACCTTTCTCTGTCGCCGAACAGCGTCCGCTCCAAAAGTTTCGGGAA
      |-----|
      313614
313664 TCAAATGGACATTGAAACGCTGTCTTCGAC
      |-----|
      313679
...
313855 TGTCCGCGGATTTTCGTCCTCGAAACGCGCGCGCTTCGATGTTGACGCAATCGCCTTCC
      |-----|
      313905
313915 TTCAGACGCCCTAAATTTGGTTTTTGGCAAAAGTTTCCGCCATTAAATCGAAACTGACGCGG
      |-----|
      313930
313975 TTTCCCTTCGATTTTCGGTAAATCGTCAGGCAGCAGCGCTTGTGGCGACCCGATGCGCCGCT
      |-----|
      313993
314035 TGCAGATTGTCGCCGCGCTCTTCGGGAAGCTCGACGACATAAGTCTGAAATGCTCCGAC
      |-----|
      314067
314095 GGGCGGTGGATTGCGGTCAAGTTTCCCA
      |-----|
      314120
...
314438 GAATCGCGCTTTGCAGAACTCGCGGGGATATACATAATGTTGTAGTCATCGTTATCG
      |-----|
      314488
      314497
314498 GCCTTCCACATAATAGCATTTGCACAACGGTGTGGCTTTCGACACGGTTTCAAACCTCG
      |-----|
      314502
314558 CCACCAATCGTCGCGAGATACGGCATATCGCACGGCATATCGCACGCAACCAAGCCAGTCGCGAG
      |-----|
      314565
314618 TCGCCAACTGCAATCGTTGGCTGCGGTACACAATGCCGAAAGCGGCGCAATGCTGCG
      |-----|
      314622
      314631
314678 ACTGCCGTG
      |-----|
      314690
...
314698 ATATCGCGACTTCTTCAGCATATTCTTCCAAATTCGGTTCGGTTCGATGGGATATGG
      |-----|
      314748
314758 CTGACCTGCGGCTGACCCCTGTGCGATGACATGATCTATCAGTGCCT
      |-----|
      314780
...
315173 TGGCCCTGAAACTCCTGACCGGACTGTGGTTCGGGTTTCGGGCAATGTCGTCTGCTTTTGA
      |-----|
      315223
315233 CACTGCTGCTCTCTTTGCGTCTGGAACGCGGCTCCGTCATCGA
      |-----|
      315256
...
315353 TTGACAATCAGTTGCGGAATTTGAAAAAGTTTAAACGCAATTGCCCAAGGATGCCA
      |-----|
      315403
315413 TCATCCGCTGATTCTTCGGAACACACCCCTCTTGTCTTATGATTGAT
      |-----|
      315446
...
315468 GCTGATTATAGATTGGCAGGCACACATCTCCCGCCGCTCCAGTCTACCGGGACCCGAC
      |-----|
      315518
```

```
315528 TCAGGTGATCTCTACCGCTTTGCCGGAACATCGAACTGTTTTCGAGGCATTTGAAAAA
      |-----|
      315588
315588 TGCCAAAGAAAAAACACACATGGTGGCTCAGGCGTTTTCATATGGCAATATATGTTGATGAC
      |-----|
      315617
...
315648 GCTGGTGTCTGCTGTACTGATGCTGTTTGGCACACAGATTTGGGTATTCGGCGCGCTGCA
      |-----|
      315650
      315659
315708 GCGGTTAAAGGAAAGGTGCGGAACGCATCGGACGAGGTGTTTCGATATTCGGGTTCGCCGA
      |-----|
      315708
315768 AGGCGGTACGCCGAATTCAAACAGGTCGCGCGTGTGTTTCAATCAATGAGGCGCAGGTT
      |-----|
      315797
315828 GAAAAATTTTATATGATGATTGGAAGGACAAGTCGCCGAGCAGACACGCACTCTCGAAAA
      |-----|
      315859
315888 ACAAATCAAAACCTGACCCCTGCTGTA
      |-----|
      315900
...
316136 CATTGGCAGGCTGTTGCTCAGCTTTCCAAACGGCATTTCTCTGTGATGAAGACGACCCGA
      |-----|
      316186
316196 TCCTGTCTCAACACTAGGCAGCAATTTGGCGCTATCGCTTGGCGGCGCAAAACAGGAGG
      |-----|
      316230
316256 AAGAAAAACGCCCTGCTTCAGATATTGCAGGAACCAACCTGATTGCGCAAGGATTACATG
      |-----|
      316262
316316 ACAGCATCGCACAGCAATTAACTTCTCTAAACCTACAGGTACAGATCTGGAACCCGCT
      |-----|
      316335
316376 TTGCCGAAAAACAAAC
      |-----|
      316400
...
316873 ACGTTGCACTCTGAAGAAAGCTTGAATGACTATAAAATTTATTCGATAGACGACCATTA
      |-----|
      316923
316933 CCCTCTTCGCGCAGCGGCATTAAAGCCCTTTTGTGCGGCCCAACACGG
      |-----|
      316960
...
316999 CCGCAGACGGCCTCTCGGGTATCAAAATGATCAGTCGCTGCAACCCGATGCTCTCTGTCG
      |-----|
      317049
317059 TTGACCTTGATATGCCCGGTATGAACGGACGCGGACACTCTCCCA
      |-----|
      317080
...
317110 TCAGCATCAATCCGACGAGCAGTGTATATGCTGACCGCTTTCGGAAGACAGGACGATTT
      |-----|
      317160
      317163
317170 TGACCGAATGATCGGCATCGGCGCGCGCTACCTGCTGAAAAAACATCAACGCCGACT
      |-----|
      317220
```

310744 TGCAACTCGATATGTCGGAAGAGCGCGGAAGAAATCGCCAAACGACGCGCGGTACGC |----|
310779
310804 CGGCGATCGCCAAACCGGCTTTGCGCGCGTGGCGATTTTCGGCGAGCTGAAAAACAACG |----|
310809
310864 GCACAATCGAGCGGCGATCGCGGATGCCGCTTTAAGTATGCTGGAGGTGACGTGCAGG |----|
310881
310924 GGCTGGACGTGAT
...
310990 GTTTGGACAATGTTGCCCGGCCATCGCGGAATCAACAGACACCATCGAAGAGCTTATCG |----|
311040
311050 AACCCTATCTCATCCAAACAGGCTTCTGCAACGACCGCGCGGCGAGGATGGCGACCG |----|
311102
311110 AAGCGCCTACCTGTCATTTTCGGGCTGCCGTCGAAATAACGCAATG
...
311285 AAAATCGCAAGACCGCAACAAAGGTATCCGCGCGGCTTGGCGGCGCGGATATTG |----|
311335
311345 TCGGTTTTGATCGCGCGCTCCACTTCAATGGCAATAGCTGCGCG
...
311490 TCAGCAAAACCATATCCAGCCTCTCCAATACGTTTTTCCAACAGATATACGGCGTTGCCG |----|
311540
311550 GATTCAACACCGACCGCGCTGACAGCCCATGTCGCCGAATCAGGCT
...
311613 GCGCGCTCGCTTCGGGATGGAACGTGATGATGCTCCAGCTTTGGCAACGACTGAA |----|
311663
311673 TCAGGTATCCACGGGTTTCGACCATCAGATGCACATCGATCGGACGCTTGCATAGGCT
311733 |----|
311735
311793 CGTCAAAATGGATCAGATCTGCACCTGCGCAATGACGCTTCCACCTCTTCGCCAAGCC |----|
311827
311853 GTGCAAAATCTGCCGATAAATGCTGGTG
...
311903 TTTTTTCTTCAATATCCTTTTATAGTGATTACAAAAATCAGGACAAGCGGAGGAG |----|
311953
311963 CCGCAGACAGTACAGATAGTACGGAACCGATTTCATTGGTCTTCAGCACCTTAGAGAAAT |
312021
312023 CGTCTCTTTGAGCTAAGCGGAGGCAACGCTGTACTGTTTTTTGTTTAATCCACTATACGT
312026 |----|
312048

312083 TCCGATTCCGCGCTTATGTCT
...
312270 CTTTGTCTTCTGCCAGCTTTCACGCGGCCACGGCTTATATCCCCCTGAAGGATTTTC |----|
312320
312330 AACCGAATCGGACATACGCCGCTCGGACTGACACAGGCTCAGCACAATGAAGTCCGTA |----|
312390 AAATCCGCGCGCTTCAAAATGGCGGACACAGGCGCGCTTTGAAGGTTATGCATTCCG |----|
312416
312450 AACACAGCGCGCTCGGTCTGCTCGTGAATCATTTCTCGGATGTTTTTAATCGGAACG |----|
312470
312510 AGCGCGCGATTATGT
...
312604 ACGCAACAGCAGCAAAATGTGGCTTTCTCTGCCTCAAAATAATCCCGGAACGCTCACA |----|
312654 312663
312664 ACGCGCGTTGTTTCGGCAGCCTGCCGCCAGTTCGAGGCAAACTCCACGCGGAACGTC |----|
312668 312718
312724 CCGAAGCGCTGCCCGCGCTGCGCCACTGCAATGCCGCCATCTGTGCG
...
312780 TCATACGGCAGCGCGAAATCTTCCAGCCAATTTTGACCGCGCCAGATAATCGTTTGA |----|
312830
312840 TCGAAGGATAAAACTGAGCCACAAATCCGAATCGGTGCGACAGGG
...
312954 TCAAAATATTCGGGCATCAGTGCCTCTGTTGGAAGTCGCTAAACCAAGCTTGGCG |----|
313004
313014 CAACGTTGAGACAGACCGCGCTTAACGGGTTTTCAATGCCTTATAGGTTTCATCGCGG |----|
313015 313067
313074 CTTTCAACGACAAATCGTCGAAATAACGATAAATTTTTCGGAGGTCTTTTCAAAGC |----|
313088 313117
313134 GTCAACAGTAAAGCAGCGCGGATTAAATCGCTTTTATCGACTTCGATCAGGCGCAATCC |----|
313136 313160
313194 TTATCCGCATATTCGTGCAGCA
...
313424 AAGCGGTTCCAATCCGATATGCCGGGTTTCGGGGAAGCAGCGCATCCAAACGCGCA |----|
313474
313484 AACGCATAGGCTTTTATCGAGAAATTCGTTCAATTCATCTCTACCTCATTTCGCATATC
313544 TTTGCGCCATCAGCGCTTCGACGGTATCGACGATTCGCTGCGTATTCGGATCGATTCGA |----|
313572

308032 AAATACAACGCGCAGCTGATTTCGTCTCCACGACCGTCA GTTCG 308022
...
308090 CTACCCAAATCATTTGAATTTGGACGGCAAGCGGATATGAACACTACTTGGCGCATACG 308140
308150 AAAGCTATCTCGAGAAAAAGCGCTAGCATAAACGCGGTTGGAAC
...
308278 CCGTCCGCGATTATCTTTTGGGCTTGCCGTTCAAAGACCGGATTTGGGTGTCGTCGCGG 308328
308338 CAGACGCACAACCATGCTGGCGCAAGGCTTCCAGCCAGTCGGCAA
...
308443 AGCGAAAAACCGCCAAAGTTACGTCTGCTTTCCACGCGACAAAGACGTTACGC 308493
308503 TGGAGCAGGATTTGATGCGCGCGACCTGACCATCAACGCGGATGGC
...
308647 CCGTCCGCTACCTCGCTACTGCGCGCTTTGCGCGCGCTTACAAGTTTGAATCGCCGAAG 308697
308707 AAACCATAAAGCTGATCGGCGAGATGGTGGAAAAACGGGAAGCGGA
...
308895 CAACGCGCGACTACCATCCGAAATCGACAGCGGCATCCATACCTTGATGACGCTTGCAA 308945
308955 CGCGCGCGCATATGGGTTGAGCTGCCGGAAGCTATGCGCGCCCTGCTGCAGCACTTG 308986
309015 GGCAAGCCAAAAACACCGTCCGACATC
...
309172 AAGTCGGACAGCTTAAAGCCAAACCATCTCTGACGTATTGAAAAAACCAGCGCTTTCA 309222
309232 GACGACCCGACGCTTTTCAGACGGCATTTGACGCTCTGCATTTGCCGACACGCAAGCGCGTC 309241
309292 TGAACCGCGAACAACACGCCCTACCC
...
309335 CGCACTGGCTGCGCTTACTCGAAGCCGCCAATCAGGCGGATTCGGGCAAAATCGCCCGCG 309375
309385 AATGCCCTCGCAGGAAAAAGCGGACCTTATCGCCGAACAAATCGACCGGCGCGGCTGG 309414
309445 CACAATCGCCCATTTCCAAAAAGCGTTTCGAGGGCGCGGAAGACAAAAACAGAAAAACATT 309450
309450

309505 AAAACGTCCAATGACGCACTTTTATAGTGATTACAAAAATCAGGACAGCGGAGAA 309507
309555 GCCCAGACAGTACAGATAGTACGGCAAGGCGAGGCAACGCTGTACTGTTTGTGTTAAT 309601
309625 CCACTATAAGTTTGTAGGACGATACCCAAATCCAAAGCTTTGCCAACGCGCCGCCATATC 309643
309685 CGCTATAATTACAG
...
309728 CATGACCCTGAAACCGATTATTATTGGCTAAATCAACAGGAAGATTATCAACGCCCTCAT 309778
309788 CCTCAACACAGTCGCGAATTTAGCGGTGGGAAATCCGCTATTGAACGAAATCCTCGA
309848 AAAATTCAATTCACGCTTTGTTACGGCGCAGGCGTTGGCGCAGCGGTGATGCAGCAAGT 309861
309908 CCGCTTCGACCCCAACGCTTATCACATCGACGACGACGACGACACACAGGCATCTG 309921
309968 CCCACACTGCATCAACCCGCTATGCGCCCTGCGCGACTATCTCTGTTGGCGGAAAC
310028 GCGCGGTAAACGCTTTTGACCGTTATCTTTTCAATGCCGCTGGAACGCGCGGACCG 310038
310088 TTCAGAGCGCATACCCGACAAAGGGAACACTATGCTGCAAAAC
...
310173 CAAACCCCTCCGACAGGAAGAACTGCTGGAACGCGCCCTCCGCCCAAAACGCTGGAC 310223 310230
310233 GACTATATCGGGCAGCATAAAGCCAAAGAACAGCTCGCCATTTTCATCCAAAGC 310235
...
310318 TCTTGCTTTTTCGGTCCCGCGGGCTGGGTAAACCACACTGGGCGCACATCATCGCCAAAG 310368
310378 AATTGGCGTAAATTTTGGCGCAAAACAGCGCCCGCTCTCGAAGCGCGCGGACCTCG 310428
310438 CCGCCCTTTTGACCAACCTTGATCCGCAACGACGTTATGTTTATCGA
...
310624 CGACCAACCGCGCGGTATGCTGACCAATCCGTTGCGGACCGGCTTCGGGATCGTCTCCC 310674
310684 GCCTCGAGTTTACGAAAAACCGGACCTTCCACCACCTGTCAGCGGTTTCGGCACAACTCC 310719

304864 TTGGCTTCGGCAAAACGGCAGGGTTTTCTCTTCGGGACTTCTTTGGCGCGAGCACCCAC
304914 |-----|
304924 GCGGTTTCGGCTGTGATGGTACGACTTCGGAAATGTGTGTTTTTCTTCAATACGTGCTCG
304977 |-----|
304980
- |
304984 CTGAATACGGCATGTGATCAGGTTTCGGGCATACCGGACATTTGCGCATCCT
304985
...
305074 CCGCTGTTTTTGGCGGCTTCGGCAAGCGAGGAAGGATGTTGAACGCTCATCGCCACG
305124 |-----|
305134 TTTTCTTTTGCCCTATTGAAGTCGGCAACCGCCTTTTTCATTTTCA
...
305272 ACTTCTGTTTCGCTGACGGTCTGTTGCTGCAAGTCTTTCAGATTCAGAGCGACATAT
305322 |-----|
305332 TCCAATTTGACCGCTTTGGGAAGCAGGTAGTCTTTTTTTGTTGCTTATAAAACTTCTGC
305392 AAATCGCTTCAGACACTTTGACTTGGCGCATGAATTCGTCGGGGTTGAAGTGTGGAA
305418 |-----|
305452 CGGATGGTGGGTTGACCTGCGTCAGCCTGATCAGCTGTTCCGCCTGCGCGTCGCGGACC
305502 |-----|
305512 AATACGCGGTTTGGAGGAGGTTTACCAATTCGCAAGGCAAACTGATCGCGGATTTCT
305572 TCGACAAACTGGTCTTCAGACATATGACGTTTGGGAAAGGTAGCGGTTTAAAGCGGTGG
305597 |-----|
305632 TCGAATTTGCGGTTTTCGTCGTGGAAATTTGGGATCGTCCACGATAATCTGCTTGATTTGT
305664 |-----|
305692 TCGGAAGAAACCGAATGCCCATCAGCT
...
305761 AACACCGCGTCGCGCGGAAGGCGCGCGCTCGGCTGTCGTTCTGTATGGCGTTGTTG
305811 |-----|
305821 ATGGAGTTCGCTGATTTTTTTCGTCGCCACTTTGGACGATGTAGTCGCACCCGGATGG
305829 |-----|
305844 |-----|
305881 GATACGCTGCTGACCCCGGAAGCGGACGAAGGTTAATGCAATCAGGCCCAAGGACTTGG
305941 |-----|
305944 GCGGCGTTCGTGTAATTTTTTCGATGGAATGGAACATATTTTAAATCGGGATATAGAATGG
...
306074 GGTTTCTTCGCTGAAGCTGGCGTAACCTTTGGCGCGCAGCTCCAATACGGGACGCTGTAT
306124 |-----|

306134 CAGGCTCGGCATTTTCGGACATCATGTTTACGGCTCCGCGCTCGAG
...
306647 TGATCGCGCGAAGCGCTCAGGCAAAATCCACCTTCATGAAAAATCCTCGGCGCGATTTGG
306697 |-----|
306707 AACAGACAGCCGCGAAGTGGCGATTCAAAACGGCGTGGGTAAATTTGGGCCAAG
306726 |-----|
306767 ACCAGTTTGCTTACG
...
306784 GATATCGCGTGTCTGGACGTGTTAATGATGGTTCACACCGAAATGTGGCGCGGATGACC
306834 |-----|
306844 GAACGTGATCGGATTTACGCCAATCCGAAGCCACCGAAGACGACTACATGAAGCCGA
306883 |-----|
306904 GAACTGGAAGCCAAAGTTTCGCCGAATACGACGGCTA
...
307214 ATATGCGAGATTTGGATTACACACCATCACCATCTATCCGGGCAACTACGACGACTATA
307264 |-----|
307274 TGCTCGCTCCGCCCAATCGCGGGAACGGCCCTGAAAGACAATGC
...
307494 TGAAGAACCAAGCTGCACCGTCAGCGTGTGGAAGTTGAAAACTGGCGAAGCGCTTGA
307544 |-----|
307554 AACCCAGTTGTTTAAAAAGCTGAACCTTATCTTGAAGGGGACAAACGCTCGCCATCAT
307598 |-----|
307614 CGGCCCCAAGCGCGCGCAATCCACCCCTGCTGAACATC
...
307682 CAGACGCTCTGCTGCGGAGCAGGAGGAGCATCAATGGGCGGAAAAAGCCGCTCGGCT
307732 |-----|
307742 ACTATCCGCAAGATCACGAAACGACTTCGATGTCGATATGGATTTGAGCGAATGGATGC
307762 |-----|
307802 GCCAATGGGCGGAGGACGCGACGACGCAACAAGTCATCCGCGGCACTTTGGGCGCTTGC
307819 |-----|
307822 |-----|
307862 TCTTCGGCAGTACGAGATGCTGCGTGAAAAAGTGAAGTTCTCTCCGCTGGTGAAAAAGGCC
307873 |-----|
307922 GTATGCT
...
307972 GATGAACCGCAACCAACATGGACATGGAAGCATCGAATCCTTGAACATGGCGCTGAA

301735 CGCGCTGGTTTAAAGTACGAAATCCTCGCCCATCAGCAGCACCCCTGCCTTCAGACGGC
301738 |-----|
301795 ATATCCAGCCCGCTAGAAATGACAGCAGCTGAGATTTTCGGCTGCCGGAAGACCGATG
301822 |-----|
301855 ATGCGCGTGCTTTCCCGCGCGGATTTCCAAATCCAAAGCGTCAGCACCAACATGTC
301908 |-----|
301915 AAACCGCGTCCGATAGCGTTTCCCGCCACGCCTTCGCATTTCAAAATCAACTCACTCAT
301931 |-----|
301975 ACGCAAAGCCTCCGCGGTTGGGTTTTTGTACGCGCGCGCTCGGGTAGACGCTGCCAC
302024 |-----|
302035 GAAAGACAGCCCCAAGGAATCAGGCAATCAGGGCAAGTCGCCCATATCGACATCGCT
302071 |-----|
302074 |-----|
302095 GGCAGGTAGTCGATAAAATAAACCTCGGAATTGATGAGGTGGACACCGAGCAGGTTTTC
302159 |-----|
302155 AAAAAGCCACAGACCTCGCGAGGTTCCAAACCCAAAGACCGCGGAGACACACCCGC
302176 |-----|
302215 CAACGTGCCGAAAAAGC
...
302307 TGTCCGCGCTGCTTTCCGTAAACCGCATCACCAGGAAGAGACAAGTTGAACGCGGCCA
302357 |-----|
302367 CGGCGATAATCAGCGTCAGGATGACATCATCCGTTTTTCCAACTCGACCGCCTCAA
302368 |-----|
302427 AATAGCTGCGGTTGCTGAACGTCCAATCGCGTACCCAAACCGTATCCCTTTGGCGCTCCG
302444 |-----|
302487 GAATCAGCTTTGCCGTCAAGGGGGGGGTTTTTGGCGGTGCGGAGTTTTTCAGCCGAGCC
302512 |-----|
302547 CCSCAACTCTTATCCAGCGGTACAGTACGCGCGCGCTTTGGATATGCGTCATCGCCA
302600 |-----|
302607 ATGAGTTGTCAACCTCGTAACACCGGTTTTTAACACGACGACCGACCGGT
...
302774 ATGTCAAACTGCGCCGGAATCAGGTCTTCAAAATTTGCCTGCCGGCATTTTGTGCGCGTAT
302824 |-----|
302834 TCCACCACTTTGGGCTCTTCAGACGGCAAAATGCCGGGTATCTGCA
...
303035 TTCAGGAGTGC CGCGGATTTCTTTCTGTGAAGCCGTTTCATAACCGACAGACGACAATC
303085 |-----|

303095 AGCGCGTTACGCCCAAGGGATTTCCGGCAATCGAAACCATCGTGATAAACGACATAAG
303112 |-----|
303113 |-----|
303155 CCGTTGGCGTTTTTTCGCCCTGAGATACCTCAAGCCTATCCAAAGCCTCTAGAGAAAAACA
303214 |-----|
303215 ACGCTACCTTAAAAATGTCTGCAAAACGTGCCGCCCGGAGCGGTTTGGGGCGGACGG
303219 |-----|
303275 CAAAAGTTTATTGTACCGTAAAAACCCAGCAGGCTCCGAACGCCCGAGTGGGGGACGG
303305 |-----|
303335 GCGGGCAGAGCGGCGGAAAAACTTGCACAACGGC
...
303402 GTTCTTGAGGAAAAACAATGAATATCCGAACCTGCTTTTGCTTTGTGCGCCATCGCCATAT
303452 |-----|
303462 CCGCCGTGCGGCTGCTACGCCAAAGAAATCAAAATCGATGCCAA
...
303699 CCGCCCGGAAAAATGCGGTCTGAAGGTTTCAGACGGCATTTTGCATTTGGCGGGTTATGC
303749 |-----|
303759 CCGCGCTTCTTTAATCAGCGCGCTTCGTACACCGCTGCGCAGCGTTCCCGCATCGAC
303779 |-----|
303819 ATATTCCAATTCGCGCCCAAGGGAATGCGCTGCGACAG
...
304074 TCGCGCTTCCATATTCGACACGTGCGGAGGCATATGCACCACCATCAGCCGTGCGCGTC
304124 |-----|
304134 GCGTGTTCATCGGCACAAATATCGCACAAATCGCCCTTCGCAAAAC
...
304255 CGCTTGTGTTGGAGCAAAATGATACGCTATCCGCTGTGCCGATTTTCGCCGAGGTTGGT
304305 |-----|
304315 AAAACCTTCAGCGCTCGATCAATCCTTGAAGGCATCTTGTTTTT
...
304396 CCGGAATATTCGCGACCGTTATTTTGTCAACAAAAGTGTCAATTTACTGACCGTGCCTGTTG
304446 |-----|
304456 TCGACCGATTGCGCTCCTTTGGTCTGTGTTGATTTTGGCGTTGAAAT
...
304525 TTGGCAGACTGCTGTTGCGCCAAAGCCTGTGTTTGAAGCGGAAGCTGTGCGCGCATATCA
304575 |-----|
304585 TCGCGCGGGTTACAGCTGTACTTCGACAATCACGGGTGCCGGCA
...

297972 ACGACATCCCGCAAGGAAGGTTGCCAAGGGCGCATATTCCTCAAAATGCTTTT
297973
297976
...
298387 CAAACGGCATTTTGAAGGCTGCCTATCAAAATATTGCAACTGTCTGGCAGGCGCATTTCC
298437
298447 ACAATGCCAACAGGTGCTTTCCGCAATGAAGAAGAGTCCGCAACCTGCGCGCTCC
298507 GTTCCGATCGGATATCGCCGCCCTTTTCCAGGCCACCGTATGAAGCAGGCGCAGACA
298521
298567 GCCGTATTCT
...
298909 CGAAACAGGCAATTCCTCCGTCATCGGCACGGATACCAATCCGAAGCCGTCGCCTGCG
298959
298969 CCGTGCCCAATATTGCTCGTTTGGGCTTTGAAAAACAGGTTGAGAT
...
299378 TTGAAAAAGGCATAAAGGGCGCGGTGCGCTTCGGCGGCAGGAATCTGAACGCCCGCG
299428
299438 CCGCGGCACGCATCAATTCGGGCGGATTATCCTTGTGAAATACC
...
299583 GCGCGGTAATGCGGTATATAACAGCTCCTTACTCAATCCGGACAATGCATGTCGCCCT
299633
299643 TCGTCCGAAGGTGCGGCGGAAGGGCGGCAGCAGCAGCTCCCGAT
...
299790 TAGCGGCAAGGCTGCCCTGCCGTCCGACATCTTCATAATCAGTCCGATGTCGCGTGG
299840
299850 AACAGTTCAAGCGCGTAGTGTCTGCTGCTGATCATATAATCGGCTCTCCGGCAAAATATGC
299867
299910 AAATGTTCCGGTA
...
300045 TTTACGGCTGCCAATAGGCTTTGTGTGCCCGGTACAACTTTCGACTCGACGGTTCGGA
300095
300105 CTGCATTCCGAATGTTCCAGTTCTGTCGGGAACCGGTCAACAATGCCACGCCCTTCA
300164
300165 TCGCCCGATACGGGCGCACGGCAAGGCAGCGATGCCGCTGTTGTCGCCGAAGCGGTGG
300169
300225 CTGAACGACAGATGGGCGGTGTTTGGCGCCAAACACGGGTGGATTGCGCTGACGCTGAA
300228
300285 CCGTGTTCGGGAAGGAAGCGGCCAGCCTTTGGTGGTTTCTCCGT

...
300475 AGGCTTCATCGATAACCAATACGTCAACGGCAGCGGGTAACAGGGTTGAACGCGCGCT
300525
300535 GCATTTTGGCGGCGCAGCTTCAGCAGTCGGTGGACGGTTTGCC
...
300679 TCGTGCGGCAAGCGCGATGTTGGGAAGATTTTCGTCTTTACGCAAAATCAGCGCCAGCA
300729
300739 GTTTGGCAACCGTTGTCGTTTTCGCCGCTTCGCCGGCCCGGGTAATCACCATAAAAGACT
300753
300799 GCAACAGTGC
...
300866 GTTTTGCCCTCGCGCCTGCGCGCTCGGGGGCGGACGTGTCGGCTGCGCCCAAGCGTTTAT
300916
300926 CTCGGCAGCAAAATCGTATTCCAATGCCACATCCTGCCCAAAAC
...
301049 GCCACCGCTCAACGGATAAACGAATGACCGTTTTCGAATGCCTGAAACAGCGGTTCGT
301099
301109 GCAGTTTGCAAGCACTTCGTCGCCGGAACCTCGCATAGTGTTCAAAAAAGGATTGCCGC
301127
301169 CCTTGGCGCGCGCTT
...
301231 CGTCTCAAGCGCTGGGGCTTCAGACGGCGGTGTTGTTTCGGCTCTTTAGCGCTTTCG
301281
301291 CCCTGTTTGGGTGCAGGTGCCGTCTTTCATGACCATCACGCTCGAAGCGCGCGCG
301351 AGTTTCGTCGTCGTGCGTTTACGACACCAGCCCGGTTCCCAATTCGTTTTCAGTTCACG
301356
301411 ATCATATCAAAAACATTCCTGCGCTTCGCACGATCGAGATTGCCGCTCGGTTTCGTCGGA
301431
301471 AGCAGGCATTGGGTT
...
301538 GTTCCCGCGCGGTGCGTCGAACGGTGTTCAGTCCGACCTTTTCCAGCATCGCGATTG
301588
301598 CCGCCTCGCGCAGCCTCTTTCACGGCTTTTTCGCGATCAGAAGCG
...
301675 GGCAGAAGATGGTGAAGTGTACACGAACCGAGATGCGGTTACGTAATTCGCCCAAA
301725 301733

291173 TGTGCCAAACGCGCATTTGGCGTAATCGTTCAAAATCGCCGCGAAGTCTTTCAGACGGCCT
291196 291205
291233 TTGTCGATCAGGAGGAGTGTGCGTAGTCGCTTCAAGCAGGCTGCGGTGCGCATATCCAAG
291260 291292
291293 GCGATTAAATCGCGCTTGGAAACTTTGCAACGGCAGCGTTAAGCGGTGGCGCATATCCAAG
291297 291326
291353 TCCAAATGTTGTGTCGTTGTCGTAAGCAG
...
291414 CAAGGCAAGGGGGCTTTTCTCCCGGAAAAAGTTTCGTTTCTGTAAAGCGCATAT
291464 291472
291474 CGCCAAACAAATTAAGCCCGCAGGAATTTCCGGATTCTTGTTCGCGTACTT
291477
...
291578 GTATCAGTTGTTTGGGCAAAATAGCCGATATTCAGTTTTCGGAAAGGAGATGCTG
291628
291638 CCGGAGAGCAATGACACCCAAATAAACCATAGCGCGCGCTTTCCAGCGACAGGGTAA
291685
291698 TGTGTCGCAAAACAGTTTGGCTTCGTAAACCCAAATCTGCGTG
...
291915 ATGTTTGATTTGGCGCTGTGTTTGACATAGGAGCTTGTGTTGGCGAGAGCTGCGC
291965
291975 GCGTTCGTTTTGTAAAAATCGTAATTCGCGCGTATTGCGTGAGT
...
292273 GAAAAACCCAGCCGTTCAACAATTTTCCGCGCGGCGCGGTATAGCGTCAAT
292323
292333 TCTTCCAATTTGCGATGATATCCGCGCTTCATGCGCGTATTTTGGCGCTTCGCGCTGC
292393 CTCAATGCGGTGTGAAAGCCTGCAACTCGGCATCGCCCTGCAAAACGTAATCCAAGCG
292425
292453 GAAATATCAATCGGCGCTTCTTGGGAAACGGAAGCAGCGCGCAGTTTTCGGAATC
292468
292513 GAGACATCGCGCGCTCTGAGTGATTTTACCCCTTGATTAAAGCAACAGGCTCGATTG
292518
292573 CCGGTTCCGTTTTTGCCGATCAACCGACGCGCTGACCGGATTTGACGGTAGCGTTGGCT
292623
292633 TTGTCAGCAGGACTTTCAACACCGCGTTGCGAGGTGAGTTTTTGATTTCAATCATACG
292693 GAAACATCGCGCGGAAAAAGCCCGTATTTTACCTGAAAGTCAAGTGCAGTGCCTGCT
292698

292753 G
...
293098 CGCAAACTGTCGCTCTATCGGATTTTCTTCAAAACAATTTTTTACCCTGCGCGCAT
293148
293158 CATCTACACCGAAGGGGTGCAAAATCAGCGGAATTCATGTAGGAT
...
293374 CGCGCGGTACGGCTGCCGTGTTGGGGACAACGCGGAGGAAATTTCCCGTCCGCTGGG
293424
293434 GCGGAAATCCGTATTTCTGCGGTGCGATTGAGTGAAGAAAA
...
293493 GCCCGTCCGAGCCTTTGTCAAGATCCGTTGAACTGGTCGACGTGAAGAGTGCATG
293552
293553 TCGTCGTCGAATTTGTTCCGGCGGTACCGGCATTTGCCAAAGATGCGGTGTTGAAAGCCATTG
293557 293555
293613 AAAACGGCAACACATCGTTACCGCCCAACAAAAAACTGCTCCCGAATACGGCAACGAAA
293627
293673 TCTTCCCGTGGCGGAAAAACAACGTCATCTCCAATTTGAAGCGCGCAGTAGCGGCG
293696 293702
293733 GTATCCCAATCATCAAGCCCTGCGGAAAGTTTGGCGGCAACACAGGATTAATCCATCG
293789
293793 CCGGATTAATTAACGGCACCCAGCAACTTCATCCTCTCCGAAATGCGCGAAAA
293794
...
293877 CCGAGCATTTGGTTATGCCGAAGCGGATCCGACCTTCGACATCGAAGCGCAACGATGCGG
293927
293937 GCCATAAAATCACCATCATGAGCGCAGTGGCATTCGGCAGCGCGAT
...
294105 TCGAGTGGCGCTCCACCTACCCCTGATTCCCGGAAAGCGCCTCTTGGCAAAACGTAACG
294155
294165 GCGTGATGAACCCCGTGGCGGTCAACGCCCGATGTTGGTGGCGAAACCTTATATTAGGGCG
294173 294188
294225 CCGGCGCGGCGCATTTGCGACCGCTTCGCCGCTGCTCCCGATCATCATCGACATCGCC
294278
294285 GCCTGTTGAAGCCGATACCGCCACCGGTACCGCATCTGGCGTTCCCAACCCGCGCAAG
294285

287324 ACCCGACCTGGGGCTCGGGTTGCGTGTGCTCAAAAGCGCGGCCCA 287314
...
287465 CCGAAATCAACACCGGGAACACCTTGTCCGATTCGTCTTTCCGTGATGAACACGCGGC 287515
|-----|
287525 TTAAGACCTTCGATTTTGGCGTTGCCAATCCGCTTATCCGCCCGAAGAAACGCGGAT 287579
|-----|
287585 TACGCCTTTTTGTGTCATCTGCTCAAAAGCCTGAAACCAAGCGGCAAGG
...
287730 TCGGGCTGCCTGCCAACCTGTTTTACGGCAGGGGCATTCCTGCCTGCATCATGTCATCG 287780
|-----|
287790 ACAAGAACACGCCCAACCGCCCAATTTGCCGAAGAGGAACAAA
...
288346 CCGCCTTCAAAGCGGACACCTGGCAAAAGTTTGCCGCGTGGCACACTCAAAAGACCTTG 288396
|-----|
288406 CCGCCATCCAACCGGGCAGGCTTATCCGGAATGGACGCGAAAGCCT
...
288487 AATACGATTTCTACCAAACTCTGACGGACTACTTGGGCGGAACCCCTGCAAGACGATGTT 288537
|-----|
288547 ATCTCATCGCCCAAGATGGCTGGAAGCGGTTAAAAACCTGGCGGAAATCACCAGAAAGAAA 288606
|-----|
288607 GCGATGAAGACGCGAACCTGACCGTCTCTTTGAGGNAACCGAACCAGCGGCAAAAGGCA 288630
|-----|
288667 AAGCCAAACCAAGCGCATCAGCAAAATAACCGCAGGAAGTCAATCGCCCGGAGCTGG 288711
|-----|
288727 TTGCCCGCCGCTACTTTTCAGCGGCATCGCCCAAGCTGGAAGAAACCAAAAGCGAGCTGG 288753
|-----|
288787 AAGCCCTAAGCCCAAGATTTGGAACCAACACATAGAAACACGCGGCGAAGAGGGTGCGC 288787
|-----|
288847 TGAACGACGATTTGGATGCAAAAGCAAACTTTCCGCCAACTTCTGAAAACCGCAT 288849
|-----|
...
289070 GCTTGGCGTTCAAGACAAATGGCTTGCCGATTTACAAGCCGAATCGAAATCGCTTGA 289120
|-----|
289130 AAACAGTATTCACGAGCTTATACGCGCTTGAACACGCTGGAAGAC
... |-----|

289348 CACCAAGACATTTAGGGATCGCGGTTCCGGTGTCCGACCGCTTAATACGTCCTCTCC 289398
|-----|
289408 ACGAACTGACCCATTTGCTCGAAGCCCATCGCAACGCCCGTTTATGGCGCATATGACA 289429
|-----|
289468 ACTTTCTCCCAAACTGGCAAGCATCAA 289440
|-----|
...
289624 CAATGCGGCTGAATGTATAGCCCGGAGCATCCGCTATCCGAAGGTACTGAACGATATGC 289674
|-----|
289684 CCTGCATCCCTTTTCGACGCTTCCATCCAGCGGCTATTGGGAAATACGTTGACCTTCAATA 289698
|-----|
289744 CGATATACAG
...
289972 GGCTTCAACCAATGCTTTGGCTTCGTTGCTCAATTCCTGTTTCGAGAAATGAGGCTCTGCT 290022
|-----|
290032 GTTCTTTGAGGGTCTGTTGCGCGCGCGGCTATTGGATTTTGGC
...
290122 TCCGCCGTTTACGGTGATGCTGCTCGGATGTTTGGGCGGGATATAGGCTCTGCT 290172
|-----|
290182 GCTGTAGTCCGATTGGGGCAAAATCGGTTGAGAGCGACGCTTTTG
...
290362 GATGCTCGGATAAATATTGATTTTTTGGTATTTTTTGTAAATGCTTCTTGGCGTTTAC 290412
|-----|
290422 GGGCTTTCTCTTTTCGGTTGCTTGTGCTTTTGTATTTCTTTTCGGGCTT
...
290711 GACCCAAAGCGCGCAAAAGTCTGATGTTTATACCGGTAAAGGCGGTGATAAGCCCGAA 290761
|-----|
290771 TTATACCGTTATTGGCAGGCAAAAGATAAGCAACCTCGCCCGGCTTCTTTATCGCTCGCA 290820
|-----|
290831 AACTGTTCTCGATTTTCAGTTTCAATCCGTTCCAAATTCCTTCTTGA
...
290924 GCCAGCTGGATAAGGTATTTTGAATTTTTTCTTTGCTTCAAGGTAAAGGCTCTTCT 290974
|-----|
290984 TGTGCTAAAATGCTTACATGCGGCTCTGAATTCGGAAAGCTGCG
...
291113 CGCTTGGTGTCTTTTGGCGCTTTGGCTTTGTGCGGATGCTGCTGGTGGCGGCTTTTCC 291163
|-----|

...
284140 TGGTTTTTTGTTCCCTATTGGGCGGTTTACAGGACTGCTGCTGGCAACGTCGCCGCTT
284200 TGCTCCAGTTTCGGTTCGTGGAACGGCATCTTGGCTGTTTGGGCGGT
...
284265 CTCGAAAGTTTTTTTCATTACGCCGAAATCGTGGGAGACGTAFCGGCCCTGTCGCCGCTTT
284325 TGGGTTATCTTTTCGCTGATGGCGTTTCGGCGAGCTGATGGGCTTTGTGCGGAATGTTGGCC
284385 GGATTGCCTTTGCCG
...
284648 TTGGCGGGCANAATCAGCGTATTGCTTTCTTTGGCAAGATTGTTGAACGCGCGCAGCTAT
284708 TGTTCCGCAATCTTCAGATTGACCGCATCCGCACCGCCTTGGGCTT
...
284795 TCGCAACAAGCGCAAGGATTCCGCTTACCTTTTGGCGGGTTGATGCGGGCGATTTC
284855 TCGGCATTTACGCAATTGACCGCAGCTGAGCCTGAGCCTTCGGATT
...
284918 CGCTGACCACTGGCAAGTTGATTGTTGTCGATTTAGGACCTTCGGATTCCGGCATACGG
284978 GCGGTTTTTTCGGTTTCAGCAGTAATTTGGCGCTGCATTGAGCGAA
...
285059 TAACGCAAAACCTTCACACCCCAAGCTCCGGCGGCTTCATCGAGGGCGGAGACGCGGTG
285119 CTGTTGATTTCGTCGGTCTTCAACGTTTGTGTCCTCAATTCCATACGCCCGATAACGGAA
285179 CGCAGCGTTCGTTTGGCAAGCTGGGTAATCGCCATAATGTAAGTTGCTCGAACCGTATGAG
285239 GCGAGTTTGGGTCGGTTACTTG
...
285311 CAGACCTGGCTGGTAGCTCTAAAGGATTCTTTTCAGCGAATGCGGTAGCGCAGCGG
285371 TCGATAAGGAATCAAAATATTCAACCGCGCTCAGGGCGGATGGAACGCCCGCAGC

285431 CTTTCGACAACGTGGACTTCCTGCTGTGGGATGACAACAAGGATTGAAAGCCGAAACA
285491 ACCACGGCTCCAGCAAGATAATGAAAAATTCATATTCCTCCGAGTGTTAAGGTTTTC
285551 CGATGATAAGAAGTTGCCCTTCCTTGGGACGATTTAGGGCGCGGTTCCCTGGTTCAAGCT
285611 CTTCTTGCCTCCGTTATTTTGAGCCT
...
285917 CGGCAACAGCGGCAACAACCATACAGTCATACATTTCCCCACAAGCGCGTCTGTTTGAC
285977 AAAATAACGCAATATCAGCAGTATAGCCGAATTTGAAAGGATAGGG
...
286057 CGAAAAACAGCAGCGCTATTTTCAGGAAATTTTAAATGCAGTCAGGCAGGACGCTTTGTC
286117 GGGACAAATCATCTATCCGCGCGCGGGATGTTTCAAGGCATTCGCGCTGACAGCGTT
286177 CGACCGGTCAAAGCCGTCAATTCCTCGGACAAAGATCCGTATCAGCGGGCAG
...
286393 TTTGACGGTACGCGCGGACAGCGCATTCGCACGCCCTTTTAGCTGGGAACGCTTTAC
286453 CGATACCGTCATCAGGCAGCTTCGCGACACACCCAAGCACCTTGTGTC
...
286896 AGCATAGTACGCGCGGCGCATGATTGAGTGCACGCCGTACGACTTTTGACGACATCG
286956 TCAAACTCAAAAACACCGCGGACATCGCGACCGCCTGAATAGATTATCGCCCAATTCG
287016 CCGAACCCCAACGACTTAAAGGGCTGATCGACGCTTACCGACTTCAACGACGAGACAAC
287076 TGGGTAAGGTAAGGAGATGATCGACCGTTTGAGCAGGCTTTGTCGGCATTTTTTAAAAAGC
287136 TCAACCTTTCTTCCAAACCAAGCGGAGACGACGATTTGTTAGTGATGCGCTACGAAATACG
287196 TGATGCGCCATTTTTCGACCGGAGTC
...
287264 CGCATTTATGGCGAAGATTATCGGAATTCAGCGCAGATTGCGCTCCAGCACCAGCGTTATG

280585

...

280697 GCGGATGAAGTCCTCGCGGTGTGGTTCTCTTCCTCCATCTCGTGATGAGACGTTCAAA
|-----|
280747

280757 AAGTTTGGTGTAGCCCATTCGGAGTAGAGCGGGAGCTGGATAAAG

...

280878 ATATCTCTCTTTCGGTAACCTTGGTTTTCGGTATGCTGAAACAGCATCTGTGCGCTATTC
|-----|
280928

280938 GGGATGCGTGCAATTGTATGCAAAAGCTGTCCGCATGACAAAATTTTC

...

281126 CCATAGCGAAAAACAATAAGACACGCTTTTATCAAAATTTATATGAATCGTTCTCA
|-----|
281176

281186 TTAACCGACAGATTCCTTTTGAGATTATCGGATTTTGGGAATAAAT

...

281619 GCGGACCATGATCGCGGATTTGGTCGGGATGTCCGGCATCATTTCTTTATAACGTTTAA
|-----|
281669

281679 TAAGTCTAAAGAAATGTTGATAATTGGCACCAGGCGGCTTTTCTG

...

281754 GTTCATCACGTCGGGCGGGTTTCGGCAAGGATTTTGAGTGCGATGTCCAAGCGTCTCG
|-----|
281804

281814 GAAATCGGGACGAGGATTTTCGATTTTGGTGTTCGGGCTGGTTTCG

281860 CGGATGCTTTGATGCAAGTCGGCGAATGCTGTGCGCGCCGCTCGCGCAGGTCGTCGCG
|-----|
281910

281920 TCGACGGAGGTGATGACGACGTAACGCAGGTTTCATGCTTTGACGGATTCGGCGAGGTT
|-----|
281934

281980. TTCGGTTCGT

...

282004 TTCGGGGCGGTGTCCACGTCCAGACGAGCGGCGGGGTGCGAGATGTCCGCCATA
|-----|
282054

282064 ATCATGAAGGTCCCGTGCCTTTTCCTGAAGCATTCGCCGATGTTGGGCGAGGAGGCTTC
|-----|
282073

282124 TCGCA

...

282187 GGAAGTTTGGCGGATATCCATTCGGGCTTTTAAATTTTCCCTGAAGGGGACGACTTG
|-----|
282237

282247 ATGGGGATCGGGCGGTTTGTCTCGGCCCTCTGAGTTTGATGCGGCTTTGGGCTCCTCG
|-----|
282300

282307 GTTTTGATTCACTCATTTGTTCTGCTTTTCGGTGTGAATGTCTTTCAAGGTGTCGGT

282367 AAGTTTCGCGCGACATTCCTCGCGGTGGGCGAGGGTTGGACAAAATCCCGGATTTCCGT
|-----|
282376

282427 CATTT

...

282437 CCGCGTAGCCGAGGGTTGATGTGGGTAAACGGGCTTAAATCCATATTCAGCTTCGAGC
|-----|
282487

282497 GCRAAGCCGTGATAGACGGACCCGCTTTTGTATACGCAGCCCCAGTGAGCGCATTTCCGT
|-----|
282544

282557 TCTCCGACATAAACCGCGGCGGTTTGGGTCTGCCGCGCTTCGATCCGCTATTCGCC
|-----| |-----|
282568 282576

282617 AATGTGCGGATGATGCTGTTTTCAAGCGCGGAACGATGTTCGCGAGCTGGTTTCGCC
|-----| |-----| |-----|
282622 282649 282661

282677 CGTTTGAATCAATCATCGTATTAACGACCAATTTGCCGGCCGCTGATAGGTAATCTGC
|-----|
282700

282737 CCGCCCCGCTCGAATTTGGACGACGGAATGTCTGTCGCGAATCAGCAGGTGTCGGGTTT
|-----| |-----|
282754 282766

282797 CCCACCAGTCCTGTGTGAACACGGCGGGTTCGACGACCCACAGTTGCTCTTCGGTG
|-----|
282832

282857 TCGGCATTCGCTCGGCATTAAGGTTTTCATCGCTTCAAAAGTCGCAGATATTCGACC
|-----|
282887

282917 AAACCTTTGTATGATTTTCATCTC

...

282981 CGGTAGATGTTGTCCAAATGTCTTGGTTTTTCAACCTTTACCTGTACGGTGGCCAGTA
|-----|
283031

283041 TAGTTGCTTTGCTGCTCGGACGGTGGTGATGTGGTGGCGCTGCG

...

283401 ATGTATCGAGAAAGCGCGGCATCAAGCCGTGGATGGATGCGGTGGCGCTTTGCC
|-----|
283451

283461 GCCTTGTCTGGCTGGTTTTTCGCGCTCGCGATACTTTTGACTCCGTTTTCGGGTTCGGCG
|-----|
283488

283521 GTGTCGCGTATGTATTGGACCC

...

283919 CCTGCTGAATGGCAGCGGTGCTGCGCATTTGCCAAACTGTTCCGAGGCGCTTTGCG
|-----|
283969

283979 CCGTGCTTATACGGCATTTACAGGCAATTTGAACGAGGATATTGGGC

277395 CCGCGCGGCGCCGAAGAGGGTGTGCGGGTTGCCGCCGTTGCCGTGAAGGTAGCGTCCGGC |-----|
277445
277455 GGCATGGAAGAGGCTTCCACGAGGCTGGCGCGCCATGCGTTCAC
...
277506 GGGATTGGTCCGCGTATGTCGGAATGGTCAGAAAGTGTAGAGCGCGCTGAGCGTTCGTG |-----|
277556
277566 GGTTTGACGCGTTTGAGAAAGCATCAGTACGTGCGGGTCTTGGATGCTTCTTTTG
277626 GGCACGCGCAGACATCAGCAGGTGGTTTTCGACCAGCGGAGCAGGTCTTCTTC |-----|
277674
277686 TCCGGTCAGGAAGTGTACGCGCAAAATTGGCGCGCTCTCGGATGCCCTGTATGGCATG
277746 GTCCGCCCGCGTCTTTGGGATGTCATGGAAAAAGCGCAAGGTAGAGGATGCTTG |-----|
277764
277806 TTTTCAAAGGACT
...
277828 GCAGAGCGGTAGGCGAGCTCGTGGCTGTGCATATCCAAGCAAGCGCGGAGCTTCGG |-----|
277878
277888 ACGACGGTGAGGATGTGCTCCACGGGATAGATGTGGAACAGGTGCTGTGGAGCAGG |-----|
277905
277948 CCGACGATTTTCCACGCGGCGAGGTAGCGTCCCAATACGCCATAGAGGTTTCAGGAAG |-----|
277950
278008 CGCAGGGTTTGGGTCACTCCGTTG
...
278101 CGCGTCCGCCCCACCATGCGCGCAGGGTTTGGGTTTCGAGCGCGGTAAATATCGTTGCGC |-----|
278151
278161 TGCTGCATGATTCGACGATTTTGAAAAATGTGTTCCGGCGCTCTGAAAAAATATCGGTG |-----|
278175
278221 TGCGCGCGCGGATTTGGTTGTTGACTTGGATGTAGTCGTCGTCATCCGAGGGTAACG |-----|
278256
278259
278281 CGCATCGCGCGGAGGAACCGCGCTTCGACGA
...
278377 AGTTCTTCGCTTTGGGCGGAGGTTCAAGCCTTCATACCCATGCTTTCGGCGGACTTGC |-----|
278427
278437 GGCTGCAAAATCGAACACGACGCGGTCTTCGGCGCGCTTGGCGGTTAAATGCAGATGATG |-----|
278475
278497 CGGATGTGGCGGAGGCGGCGGTAGCCGTGCGAAA
...

278537 CCGCTTCGGCACGCGTCAAAATCCGCTGTTTGGAGCAGTCCGGTCGCTCGCCAAAGC |-----|
278587
278597 CTTGGCGCTTCGCTATCCAAAGCAGGGTGTGGATATCGCGCAGACC
...
278687 CTTGCGATTGGCGTGGCGGTGTTCCATCTCCACCAGTTTGGCTCGACAACGCCGCCA |-----|
278737
278747 CATTTGCGCTCGCGTTTCATTTTTCGCTAGTTTATCCGCCGTTTGGCGGTTGCCAAACA
278807 AAAACCTAGCCTCTAAAACGCTGTGCCCCCGTAATATCATTTGGCGACGCTTTCACATA |-----|
278824
278867 GTTCATCAACGCTGCCGCTTTTACAGACGGCATCAGTTTGCAGTCCACAGGTTTGA |-----|
278874
278927 CAA
...
279352 TGATACGCCGCCACGCCCAAGCGCGGCCATTTTTCGATATTCACAGCGCATGTTC |-----|
279402
279412 CAGCGTTTGGCTCGACTGCCGACACATAAGTCTCTGTCCAAACCGCATTCGCCGCGC |-----|
279413
...
280017 TGTTTTCCAAACAAATCGCGGCTGACGTAATCTCTGTTGCGCTTCTGCTGGCGATGGCG |-----|
280067
280077 GCAAGCAGGCTTCGTGTTTTTCTCTGTTTGGTCAAAATCGCAGCGATGATTCTCTCG |-----|
280121
280137 GTGACTCGCCAATCAGAAAGCTTGGCCAGTCTTGCAGGTTTCGAGACCTTCGAGGAAC
280197 AGAATGCGCTCGATCAAAATCGTCGCGCAGCTTTTCATTTCAACGATGGAGTCTTTGAAGAA |-----|
280214
280235
280257 GTTTCGCCCAGTTCTTCAAAGGCCCGAGTTTTC
...
280323 GTTACACGACGACAAGCCTAAGTTTTTGTTCAGTCCGGGATACCAACAGGTTCGCTTTC |-----|
280373
280383 ATACGGACTCCTTTTATTCGAAATACGGATTACAGTTGGCTTTGGT
...
280464 AGCTGTGTTCCAGCAGTGGCGTGTCTCTCTCGGTATCTTTTCAGTTGGCGGACCATC |-----|
280514
280524 AAATCCGGCTGACATAGTCTTCGCCCTCTTCGCACAGTTTGTATGCTTTTTTCAAGCG |-----|
280580
280584 TCGCCGACTTCTGTATTTCGGTTTTTCAGGTTCGGCTTTTCAGGCAGGAACACAGT -|

273457 CGAGCGCCGACCGCGACGTGGAACACTGGCTGAACACCTATTATATGCCGCGATAAGGT 273447
|-----|
273458
273461
...
273561 GTATCTTTGTAACACTGGACGGCATCCACATTGACGGCTTGGTCGNATATCAGCGATTGG |-----|
273611
273621 GCGAAGACTATTTCAACTTCGCCGCCGAAATCATGGAATCATGGAAGCGAAGCGACGCGCA
273681 TCCGTTTCAACATGGGGACAGGGTTGCCGTCCGGGTCCGCCGTGCCGATTGTTGATGACG |-----|
273716
273741 GAAAAATCGATTTTGTCTCTGATTGCCGGGG |-----|
...
274028 CCGCCCGATTTTTGATATGGCGCGCGCGCATCATTTACGGCCAATCGAAACCCCGGTA |-----|
274078
274088 TTGAAGATTGCGTTTATTTCACCGCGGTTTTAAAGCGCGCGGTATTTCGGCAGACGGG
|-----|
274148 GCGCAACGGCGTTTCAGACGGCATTTTCATTCTTTTCGGCGTGTCCGTCCGAAATTGCTTTG 274156
274208 CCGG
...
274304 AAAGCGTTTCAAATCAGCGGTGGTAGTTTCGGTGTCTTTTAGTGATTGTAACGTCTGTGA |-----|
274354
274364 ACGTGCGATTTCGTCATACCTTCGCGAAGTGATTTCACAAATTCGG
...
274658 TCGATTTCCGCCCGTGCTTCTTCGTCGCCGCAACATACCGCCGAGCATCGACGTGTAC |-----|
274708
274718 GCGCTGCGCGAGGGCTTTGGCGATGTCGCCGGAGAGCGATGCCGCCGTCCGCAATC |-----|
274738
274744
274778 AGCGGAACGCCCGTGCCTTTGAGGSGCTTCGGCAACGTTGTGAATGGCGTCAAGTTGCGGC |-----|
274783
274838 ACGCCGACACCTGCCACGATACGGGTGGTGCAAATCGATCCCGGACCGCATACCGACTTG |-----|
274898 ACGCATCCGCCCGCGCGGACCAAAATCCAAAGCGGCTTTGGCAGTGGCGCATGTGCGG |-----|
274915
274958 CCGATGACTTGGATGTGGGGATGGTTTTCCTTTGACCCAAACGACGCGGTCTCATCACGCTT |-----|
275018 TGGCTGTGCCCGTGGGGGTATCCGACGACAAATCACGTCCACGCGCGGCGCTCAACCAAGCT |-----|
275041
|-----|

275078 TTGACGGCGCTCTTCGGTGTTCGCCGCCCGGTGCCGACTGCCGCAACGACGCGACAGCGCCT 275095
275138 TCGAGTCTTTGT
...
275257 CATCAGTTCCGGCGCTTCGTCTATGCTTGTGCCCTTCGGGGACGGTTAACCAAGACGTTCGCG |-----|
275307
275317 CGGGGTCAATATGGCGGAACGGGCAAAATCAGCGGGTTTTCAAACGCAAGTCGCGGTTT
275377 GGTACGATCCGACGACGCTTTTCCGCTTTTCAACGACGGGACGCGGACATTTTTCGCTTT |-----|
275389
275437 GCGTGCACGCGCATTTCCAAGACTTCGCGGATGAGCGTTCTCGGTGCAACGGTTACGGG |-----|
275471
275497 GTCTTTGACCACGCGGCTTTTCGTGGCGTTTCACCTTTGGAATGGCGCGCTGCATTTTC |-----|
275520
275557 GGGCGCATGTTTTTATGG
...
275651 AGCAGGGGAGGTTGAGTGTGATTTTCGGGTGAGCTTGGTTGAAGTTTAACTGCTCGC |-----|
275701
275711 GGCAGCAGGTCGAATGTGCGGGAACCAACAAATCGTCGAAGTATAGGCTTTTCT |-----|
275746
275771 ACGATACGCATAATGCTCGGTCTTTTCAGTTT
...
276606 CCCGCTTTCAGACGGCATACGGCGTGGCGGCGGAATGCCGTCTGAAGCGCGCGCTTATGA |-----|
276656
276666 TAATTGTTCCAGCAGCGTCTGCTTCAATGAGGACTGGATTTTGGGTTTTTTCAGTCTAG |-----|
276679
276726 GCTAAAAAC
...
276742 CTGTCTTCGGCGGCTTCATCCAGCGTGGAATTTTGGGTAGCGCAGGCTGACGCTGTGG |-----|
276792
276802 GCGAAAAAGACTTCGCCCATATCGCGAGCAGGAAAGGCGGTTGA
...
277159 CGGCAAAACAGCGCGGCCCTTGGGCATGAAAACCATAACTTGAAGCTGCTGCTTTTG |-----|
277209
277219 AACAGATGCGGCTGCGGACGATGGGGTTTTCAAAATCGTGGACTAGTTGGCGCGCTGC |-----|
277236
277279 CACAGGATTTCCG
...
|-----|

270730 CTACCAATCAAGAACTGCCGCCATTTTCCAGGGCATATCGCCCGTCGCGTTTCGG
|-----|
27070
270790 TATCGGGTTCGGCTTCGGGCGAGCATGTCGAACAAT
...
270854 CCTTACAAGCTTAGGAAGAGATATAGTGGATTAAACAAAATCAGGACAGGCGACGAAG
|-----|
270904
270914 CCGCAGATAGTACAGATAGTACGAACCGATTTCATTGGTGCTTCAGCACCTTAGAGNAT
|-----|
270972
270974 CGTTCTCTTTGAGCTAAGGCGAGGCAACGCCGTACTGTGTTTTTGTAAATCCACTATAAAA
|-----|
270999
271034 TATACCACAAACATACCATAAAAAACGGTAACGGATAGTTGTAAGCGGTTGATGACGA
|-----|
271089
271094 TTATAGACAATAACCGGTTTTTCCCAATGAATTTATTGTTTCAATAATAAAA
|-----|
271094
...
271512 GGGCTTTGCTTTATTGCTGTTTTTGGGTACAGTGCTCTGCTGCAAGAAGTCGCTATTA
|-----|
271562
271572 TATAGTTCAATAAGAGAATGTCAACAGTCCAAATGAATAAAATAT
271618 TAAATCTTTAAATTTACGGGAAAAGACCCGTTTTTAAAGTCGTGAAAACAGCGGTATGA
|-----|
271668
271678 ACATCCTTTGCCCAGTCGGGAATGGATAATCGAGCTGCTTGAACGT
...
271747 AGCTTTGGTAGCGCAATGTGCGATTAAAGGAAGAGATACGAATTTTTCGAACGTCGCT
|-----|
271797
271800
271807 GAAGCGCATGGCGGGGACGTCAGGTTTAAATCAACCGTCGGGGCGCGTTTGC GCGG
|-----|
271810
...
271921 TTTCGCGGTGCGGCTCAGCCCGCCAAAGACGGTGATTTTGTGTACGAACGCCAGAT
|-----|
271971
271981 GCGGGCATTATGACGCGCATATTGTCACTGTGTCCTGCGGCGATGGACCGTAGGGG
|-----|
271997
272041 CCGCCGCGAAGG
...
272072 GAAGCGCGCAAGCAAGTGGTGGCGCGTTTCTATATGATAGGGCGGTGGCGATTTG
|-----|
272122

272132 GAGCCGGAAGACAAAGCGTCTGAACCAAGCATCGATTGGAACGGACGGCGTGGCGGT
272145
272192 TTCAAACTGAATCCGGTCAGGTCATCGTCGGCGAAATTTAGGTTTATCCTTGAGCAAAAC
|-----|
272216
272252 CGGCCGCGAGTGGCAAAATCATCGAAGTTTGGCCGATATATGCCGACAGCGGCATGGAG
|-----|
272285
272312 ATTGAAATTCGCGTGCACAAGCATATTT
...
272355 GTGAAGCGTGTGCCAAAGCCGCAAAAAATTTCCGACCATGTACGCAAAAGCGATTGTA
|-----|
272405
272415 AAGCCCGCTCGATTTCGGCGACCTGCTTTGGTAAAGATAGAGCGGGAACGGCTCGAG
|-----|
272449
272475 ATTTTGACGATGCGGTGTTTTGCCGAGAAAAATCGGACGCAATTAACCGTCTGTCGTGGCGA
|-----|
272480
272530
272535 TTGCCGATGTCAGCCATTATGTCCGCCCGCATGACGCTATCGACACGAGGCTCAGGAAC
|-----|
272567
272595 GCAGCACAGTGTGTTACTTCCGCGCGCGGTGATTCCCATGTTGCCGGAACCTGTCCA
|-----|
272685
272655 ACGGCATCTGTCGCTCAATCCTCATGTGAGGCGTTTGTGTGTGGTGTGGATATGTTA
|-----|
272715 TCACCTACGCGGGCAATATCAAGAA
...
272833 CAAAACCAATCGACACGCTTTTACAAACTCTTCAAAATCCTTCAGAAAAAGCGTTTCGA
|-----|
272883
272893 ACGCGGGCGGTGGAGTTTGACAGCATCGAAACCAAAATGCTTTTCGACGACACACGTA
|-----|
272939
272953 AATTGAAAAATCGTCCCGCTTGTCCGCAACGATGCCCCACAAAGCTGATTGAAGAAATGAT
|-----|
272981
273013 GTTGGCGCAACGCTTTTGGCGCAGCGGATTTTCTGTTGAAAAACAAGCATACCGCATTTGT
|-----|
273023
273073 CCGCAA
...
273105 TCGCCGCTTTCGGGAGCAGCTCGGTCTGTTGGGGCTTCAACTTGGGGGGGGGACAAACC
|-----|
273155
273165 CGTCGCGGAAGACTATGCGCGCTTCCGCGACAGTTCAAAAGCAGCGCCGATGCCG
|-----|
273166
...
273397 CACGCCAAAAAAGCTGGCAGGCTTTGGGCGTGCATACCTCTTCTGTGAGCGCGCTGC

268343 TGGCGGCATCATCGGCAGTCTGCTCGCACAGGCGGTGCCGGTTTT 268333
...
268433 ATGTCATAAAGAANTCGCAGGCATTGGCGCAGGGCTGTGGCAGGGGAATCGCCCGG 268483
|-----|
268493 CGCAAGGTGGCTCGCTAACCGGATAACTGAAAGTATGTAAGAAGATATAGTGGATTAAAC
|-----|
268553 AAAAAACAGTACGCGCTTGCTCGCCTTAGCTCAAGAGACGATTCTCTAAGGTGCTGA 268592
|-----|
268613 AGCACTAAGTGAATCGGTTCCGTACTATCCGTACT 268592
...
268771 TTATCTCTTACAGTTTTTTGAAAAACATTAATTATGAAACAAATCCGCAACATCGCCATCA 268821 268830
|-----|
268831 TCGCCACGTCGACCGCCGCAACCGCAACACATTGGTCGACCAACTGCTCGGCCAATCCGGCA 268835
|-----|
268891 CATTCGCGCCCAACCGCAGCTTGACGAGCGGTGATGGACACGACCTTGAAAGAG 268935
|-----|
268951 AACCGGCATCACCATCTCGCCCAAAACACCGCCATCGATTACGAAGGTACCCATCA 269013
|-----|
269011 ATATCTGTCGACACGCGCGGACACGCCGACCTTCGGCGCGAAGTAGAGCGAGTTTGGGGA 269013
|-----|
269071 TGGTGGACTCGCTCGTCTTGTGGTGACGCGCAGGAAGCCCGATGCCGCAACCCGTT 269082
|-----|
269131 TCGTGACCAAAAAGCCTTGGCTTGGGGCTGAACCGATTGCTGCATCATCAATAAATCG 269172
|-----|
269191 ACAAAACGTCGCCCGCTCCGAGCTGGGTATCGACCA 269340
|-----|
269290 ATGCTTCGGTCTGTCCGGTTTCGCCAAATTGGAAGAAACCGACGAGAGCAACGACATGC 269340
|-----|
269350 GTCCGCTGTTGATACATCTTAAATATACGCTGCACCGAGCGGCGCGGAGCAAA 269408
|-----|
269410 CGCTGCAACTGCAAAATTTCCCAACTCGACTACGACAACATACACCGCGCGCTCG 269413
|-----|
...
269485 GAGGTATCAAGCCCGTCAAGTTGTTGCCGTCATGAACACGATCAACAAATCGCCCAAG 269535
|-----|
269545 GCCGCATCAACAGCTTTTGGGTTTCAAGGTTTAGAACCGTGGCGCTTGAAGAAGCCG 269545
|-----|
269605 AAGCCGCGACATCGTGATTATTTCCGGTATTGAAGACATCGGCATCGGGGTAACCATCA 269610
|-----|

269665 CCGACAAAGACACCCCAAGGCTGCCGATGTTGAGCGTGAGCAACCGACGCTGACGA 269714 269720
|-----|
269725 TGGACTTTATGGTCAACACACGAGCCGTTGGCAGGTACGGAAGCAAAATTCGTAACCAAGCC
269725
|-----|
269785 GCCAAATCCGCGACCGCTGCAAAAAGAAATTGCTGACCAACGTCGCCCTGCGCGTGAAG 269823 269826
|-----|
269845 ATACCGCCGATGCCGACGCTGTTCCGCGTATCCGGGCG 269826
...
269974 TCGACGCTCAAAAATGCGAACCCTGATGAAAACCTGACCGTGGAGGTACCCGACGACAACC 270024
|-----|
270034 AAGCGCGGTAATGGAAGAACTCGCGCGCCGCTGGCGAACTGACTAATATGGAAGCG 270090
|-----|
270094 ACGGCAACGAGCAGCAGCCGCTCGAATPACCATATTCACGCGCGCGCTTGTAT 270095
|-----|
...
270157 GCGAATTTATGACCTGACGCGGGGTGCGGGTGTATGAGCCACGTTGTCGACGATTACG 270207
|-----|
270217 CGCCCGTCAACCCGATATGCTGGCGCCCAACGCGCTGCTGGTGTCCCAAGACGAG 270332
|-----|
270337 CCAACGACAAAATCTAGGAAGGTATGATTATCGGCATCCACATCGCGGACAACGATTGG 270387
270337 270344
|-----|
270397 TGTCAACCGCTCAAAAGGCAAAAACCTTACCAACATCCGTGCCAGCGGTACCGACGAAG 270485
|-----|
270457 CCCTTCGCTGACACCGCGATTAAAGTACGCTGGAAGGTGCGGTGCGAGTTTATCGAGC 270513
270485
|-----|
270517 ATGATGAGCTGTAGAAATTCACGCCCAATCCATCCGTCTGCGCAAGCGTTACTTGAGCG 270562
270518
|-----|
270577 AATTGGAACGCGCCGCCATTTCAAAAAGCTAGATTGATGTTTACTGGATTAAATGTTAA 270583
270583
|-----|
270637 AT 270637
|-----|
...
270670 GGCATTTTTTATTCGAGCGGCTTTGCGGCTTCTTGAAGCTGTTTCAGACGCGGTTTTTC 270720
|-----|

264917 TGCAGCAGCCCCGTATCCAAAGTAAAAATACCGATATTGAGGTTTTCTCCGGCGATAAGG |-----|
264967
264977 TCGGTAATTACCATATCTTCCGCAGCAAGGCTGCTGGCAAAACCGTGCAATCCCCGGTGGCTG
265037 CCGGCAATCCGGTGCAGGCTTCTTTGAGGGTTTCGGTTTTCGCCAAAGCCGTTTCA |-----|
265053
265097 CTGCCATTTTCA
...
265196 AGTCCGTCCAATGCAAAATACCGGTGTTCCGGCCGGATTGCTTCTTTTGGCGTTTCG |-----|
265246
265256 CCGAACCAAGCCAGTTTTTCTGTGAGCCGCCACCACTTCGCCTATGA
...
265372 TTAGCCGCAGCGTCCTTGCGTGATGACGGCTGCCGGCTGTCGGGCGAGCGTCCGTG |-----|
265422
265432 CTGTTGCAGCCGTTCCGCAATCAGCGCGGCTTTGACGCGCACCCATA
...
265592 AGCGGATGCCGSCATAGCGGTGCGCGACGCGGCGCGTAAATGCCGGGACGACCGAA |-----|
265642
265652 AACGGAATCTGATGCGGTGCCAAGGTTTCCAATTCTTCGCCGCCG
...
265875 GCCGTCGTAAAGCACCACGCTGCGCTGCTGGATTTCCTTCAGCCCTTTGAGCGTCAGCA |-----|
265925
265935 GCCCCCATCGCGCGGACCGCGCCGACCAAGAGGAGAGCGCGCTTGATCATTTTGAC |-----|
265942
265995 --|GACTTTGTTCCAAATTGGCCTGCCAATTCCTCGTTCGCAAGGGTGTTTTGCCGGTTTTGA |-----|
265997
266055 CGAGGCGGGGAACGCTCCGTTAAACTGCTTTTCCCAAAAGCGGCGGCTTCGGTAACGG |-----|
266067
266115 ATTTCAGTTTGCCCTTGACGGCATCGCGCACCTTCTGAAATTTCCGCCATATCGCCA |-----|
266167
266175 AAGACGGCGCAGCAGGGCTTCAGCCTTTCAGCAGCAGTTCGGGCGA
...
266330 GGCTTTGGCAGCTTTCGCAAGATGTAACACCGCCGTTGAGGGCTTGGTCGTGCTT |-----|
266380
266390 GCCGAATGATGAGGAAAACCGTGGGATGTGTTCCGCACGAAATT
...
266504 GCNAACCTGACCTCTCGGCCCGCCTTCACGACAGCGCTCATTTTGGTGCAGCGACTGG |-----|
266554

266564 CCGCCGCTACGACCAATACGGGCGCGCGGAGGTTGCGGAAAA
...
267034 CCGGTTATCATCGGAGGACGAGGTTTTCTTTTGAACACGCGGTATCGGGCAGCGCG |-----|
267084
267094 -----|GCATGTGACTCGGAAAAAACGGCAGCAGTTTGAAAAAGTTGAGATGGATTT
267098
...
267200 TGGTGTGAAGGTGTTTTTCATCGGGCGCAAAAAAGCCCCGTGCCGACTTGAAGACGACAGG |-----|
267250
267260 CGGCCAAAAGCGCAACGACCATCAGTCGCCCAAGGGTTGTGTCCGGGTTGACGTGTT |-----|
267273
267320 CGGGAATAC
...
267340 TACTCTCTGATACCCGACCATCCGCGACGAAACGGGAGAAACGGCGGTATCGCTGCC |-----|
267390
267400 CAAATGCCCCAGCAGAGCGGGAATACGAGCAGGAAAGGATAAACAGCCGACGCGGTG
...
267460 TGCCATTGCAGCATATCGCGCGGTTTTTCGCGCTATACCAACATAAAGGTAGGATGCG |-----|
267474
267520 GCAAGCAGCC
...
267820 GCGCAATGTATCTCGGCTCGGCAAAAGTGTGGCGGGCTTCAATCAGGAAACGCTCCG |-----|
267870
267880 TTTCGCGTCATTCGCGGTTTTTCCGGAATATGTTGGACACGCGCAGGTTTGACCAAG |-----|
267938
267940 CGTCAAGGCATAAACGCTTGGACGCGGTTGCGGCTTGGGTACGGATAAGCGCGGCTC |-----|
267943 267952
268000 GGAACGCTTGC CGGAATTTTGACGGAACACACGCAAGCCGCTCTTTTGGATGCGG |-----|
268002
...
268078 ACCGATCGGAAACCCGAAATCTGGCACGCGGTGTAAAAACCTGATTTTAAAGCCACAC |-----|
268128
268138 CCGGCCGAAGCCGCGCCTGCTTGGAAAGCACCGTTGCACAGGTTTCAGCGGATCGGACG |-----|
268164
268198 GCGGCAGTGAGGAAGATAGGGG
...
268283 AAATCTATGTCAACGAAAGCGCACGCGGGATTGGCAACGGCGGCGAGTGGCGAGGTAT |-----|

```
...
262233 TTTGAAGGTTGATCGGACTTTGGCGATGTCGTTTCATCTTCAAGGATTCGCGCGCGGCCT
|-----| |-----|
262255 262261
262293 CTTGTTCCAGGTGCGGACATCTAAACGCTTTTCAATTTCCCGGACTTTTGCCGGCAGG
|-----|
262318
262353 TTTGCGTGCCGTGCTTGAGCAGGTATTTGCGCGGGTGTGAGCGGACGTTTCGTCAAACC
|-----|
262398
262413 AGCAACCGTGGCTTCAGATGTTTTTTCGGGGGAGCGGGGAATTTTATCGACAAAA
|-----|
262473 GGTGCGCGCGGAACATCATGTCGCGATCCAGCGGAGGTTGCGCGCTCTCCGG
262474
...
262571 GACACTTCGGCGGTCAGCCCATTCGGTTCGATCGGACGGTTTGTCCGACGCGACCGAA
|-----|
262621
262631 CCGCGTTGATGCGCCCTGATAGCCTCGGAAATCATCGGCCTTGTCGGCATCTTGCGG
|
262690
262691 ACGACCACTTCACGGGGAATAAATACTCTCGCGGTGCGGTGACTTCGTCGCGCCCCC
|-----|
262717
262751 GGCAGGGTTTCCAAAATGGACA
|-----|
262909
...
262775 AAGGTTTCGCTTTATACCAAGGCATATTCGCCCGGGGTAACAATGTTGTCGCCCAAG
|-----|
262825
262835 AGTGGGACACCGGTACGAATGCGGCTCTTCAAACCGAGCTGTTCCGGCAAGTCGGCG
|-----|
262895 TATGCTCCACAAATGCGGTTGAATTTGTCTTCGTGTATCCAGCAGTCCATTTTGTG
262909
262955 ACCGCCACCACATATCGCGGACGTCAGTTGGGGAGGATGGGGAATGGCGTTTGGTC
|-----|
263004
263015 TGGCGCAGAGCTGCAAGGCTGCGCGCCGAATCCAGTTGGGAT
|-----|
...
263093 CTTGCGCCCGTAACCATATTTGCGGTGATTTGTTGTCGCCCGGTGTCGGCGATGATG
|-----|
263143
263153 AATTTGCTTTCGGCTGGAAAAATAGCGGTATGCCACATCGATCG
|-----|
...
263284 TTCAAAGTTTTGATTTGGTCGACACAGGGCTTTGCTGTCTGTAGAGCAGTCTGCGAT
|-----|
263334
263344 TAGGGTGTCTTTGCGGTATCGACGCTGCGCGCGGTAAATGAAGCGGAGCGGGTTTGGTG
|-----|
263365
263404 TTGTGCGGTCAATTTT
```

```
...
263455 TAGAATAGCCTTCTTTTTTTCGTTTTTCCATTGCCGCTCGCTGCTGATCTGCCAGC
|-----|
263505
263515 CGGTCGCGGTGCGTTTCGGAAATGTCGGCAACCGCTGTTCCTCTGA
|-----|
...
263563 ATCTCCGTGCGGTGAGCGGTCTTCTTACCGGGCAGGTGCAGCTGATGTCGCCGACG
263613
263623 GTGCGGAAGCGGACATCAAGGATTTTCGAGAGTTTCAGACGCAATTTTCGGGGTCAGCGGC
|---|
263680
263683 GTTACGGGGACAGCAGCCCCCTCGCTCGACCACCTTCGCGCCTGTGGCTGTAATAAATC
--|
263685
263743 GCGGCAGCTCGAGGTTTTTCGCGGGCATGATTTTCGCCATGTGTCAGATGTCGATTCGTCAGTTG
|-----|
263766
263803 GAAATCGGAAGACGCGCATATTTTCGCCCTTATGACAGCCTGGTGTGTACAGCGACCA
|-----|
263854
263863 AGCTCGGGGGTTGCGCCTTCGGATCCCATTTGCCGAACCTGTCGCGGAACGAGAAAAATC
|-----|
263870
263923 CGT
...
264301 GCCACTTCGCGGATAATGTGATGATCGGATTTCAGATTCCAACCACTGAGTTGACAGCTGTC
|-----|
264351
264361 GGTTTCGTTTTTCGTCATACCATATTCCTTATTTCTTCTGTGTTGATC
|-----|
...
264497 GGCAGGTCAGGGTTCGACGCTATATCTGGGAAAGCCTTGCCGGTACAAATCGTTGTAA
|-----|
264547
264557 GGCACATTTGTCGAAGGATGTATGCCCCACACGTCGTGTTCCGACC
|-----|
...
264640 GGCATCATATTCGCAAAACGGCAGTTCCTCGGTGTCGGGACTGTTTCGGCGCTTCCCC
|-----|
264690
264700 GGTAAAGCCAAGCATCCCGCCCGCAATGGCGGTGTCAGCGGTTTCGGTTTTTCGGATGCG
|-----|
264760 GCAGCATTCGCGCGCTTCAACGCTGTCGTAAGGCAAAACCTGCTTTGCTTTCCAC
264781
264820 ATAATGAGGGCATCTTCCTGATTCGGTTGGAACGCTTGAATTTGCATATCGCGATAAAC
|-----|
264852
264880 ACGTTCAATCCTATCCAACAGGTTTAAA
|-----|
...
264885
```

259292
259351 CCGCGGTTTGGCGTTCGCTCGCGTTGCCGAAGTCGCGCTGTGTGGTTACCAACGCTTCG
259353 |----| |----| |----| |----|
259411 GCGGCCTTQA 259365
...
259504 CCGCCCGCCAAACATTGAATCCGGCAAGCTGTCCGTTGCCGTCTGAATGGCGCAAAA
259554 |----| |----| |----| |----|
259564 TCCAAGTCGTTGCCGTAGCAGTCCACATCATTTAAGGCGGGATGACGACTGCGGTTTG
259569 |----| |----| |----| |----|
259624 AATTTGCGCGCAGATAGGTTTTGCCCAAAATCGGCTCGTCTTCTTGAAGGAAGTCGTCA
259677 |----| |----| |----| |----|
259684 GAACCTTTGAACCTTTTTCGCGTCCACCACACATCCAGATAACCGCGG
...
259798 TCGATCGGTTGGACGTGCAAGCACGTTGGGTTCAATATCCGCGCGGTGGGATGGAA
259848 |----| |----| |----| |----|
259858 TCCAAACCCAGTTTGTGCAAGAGCGGTGCATCGTCTGCAACTTGGCTTTCGGCACGCG
259888 |----| |----| |----| |----|
259918 TGAATTGGAAGTTTCCGCGTTGGT
...
260050 AATTTCAAGGTCGAGTTTTTGCTCGGCGGTTTCGGCGGGATGTCGCGGTGCTCTGC
260100 |----| |----| |----| |----|
260110 TCATACATACCGTGAAGCGGATGAGTTGGAAGTTGCGCTTTGAAGCGCGCGTGAGC
260145 |----| |----| |----| |----|
260170 GGGTCTTTCAATCGTCCAAATCGTCCGCGTAAAAAATGCTTTCGGTTTTCAGACGT
260181 |----| |----| |----| |----|
260230 TCGTTGTCGATAGCGGTTTTTCTTGCCACGCCAAACCTTTTGTCTTGGTC
260230
...
260288 TCATTTTGTGTTCTCCCGATTATATTAAATCAATAACATCAGCGTGATAGGTTTTC
260338 |----| |----| |----| |----|
260348 TTCGCGCAGCATATCCAATATCTTCTGCGCGCTCTTCTGCTCCAAA
...
260406 ATAATCACATCCAGCAGCGGCTTCCACGCTCTTTTGCCATTTTTCGCGCATGCGCGCAC
260456 |----| |----| |----| |----|
260466 ACATAGATATGCGGCTTCTCTGCAGCATTGCCAAAGTCCTTCG
...
262173 CTATCAGGATAAACGAACCGCGCGGTGTTTCCCGCATAGGCGTTGCGGTACAGGGTT
262214 |----| |----| |----| |----|

260676 CCTTCCGCAATTTCTGCGGCACGTTGTTGSAAGCGGAAACGCGTGCACGCCGCGTG
260726 |----| |----| |----| |----|
260736 CCCAGCGGATCATACAATCGGGCTTCGGGCTGTCTTCGGGCAGCGCTGAACCGCTGCTTG
260789 |----| |----| |----| |----|
260796 CGTTCCACAAAACAGCGCACGTCGCGTCTCTTCCAGCCGCTCGGCAA
...
261021 CGGTGCAGCACATCGGCAATCGGCGTGTTCACAAAAGCCTTGCAAAACGCGTTGTTCG
261071 |----| |----| |----| |----|
261081 GCAGCAATACGGTCGAGTTCTCATCATCGCGCAACGGGGCATAGC
...
261219 GTTGCTGATCGATGCCAGCAGGCTTAGGATTTCCCTGACCAGTGCAGGATCTGTCA
261269 |----| |----| |----| |----|
261279 AACCAAAACGCGAGCGCTGCCCGGGAGGTAGTGCAAAATCCGAACCGCTCAAAATCGATT
261284 |----| |----| |----| |----|
261339 TCGATGTGGCGCA
...
261412 GGTCTGCTTGCAGTACCTGCCATCCCGTGCCGTCTGAAGCCCGGGGCGCTTGTCT
261462 |----| |----| |----| |----|
261472 GCGGCGGGCGGCTTGCCCGGTTTTTTTGGGCTTCTTCTTTTAAGAGTGCAGCGATATTAT
261521 |----| |----| |----| |----|
261532 CTGTCCATCCGTCTGCGGCGGCGCAAAATCCCAATCCGATCAACGCGTTCGAGCAGGC
261589 |----| |----| |----| |----|
261592 GTTTTGGCCCAATCTTCAAAACGTTTGTGCAAAATCTTTCGCCCGCGCAGAAATCG
261613 |----| |----| |----| |----|
261652 GATAGGAGCTGTGCCCAACCCAGTACGGCAAAATTGGAGTTTGTCCAATTTCCGGGCTT
261662 |----| |----| |----| |----|
261712 TTTTGC
...
261746 TTCTTCCGGCGGTTTCGCTTTCGCTTGGGTGAGGTAACCAAGCAGCAGCGGCTTCGC
261796 |----| |----| |----| |----|
261806 GCGGATGTTTTTTCGCTTATAGTCTTTCAGTTTCGGCGGACTGACTTGGATGCGCG
261806 |----| |----| |----| |----|
...
262113 TGCATATCGTGTCTCACTCATATAAATCATACCTGCGCAACAGTCCGCTATGTCTCGT
262163 |----| |----| |----| |----|
262173 CTATCAGGATAAACGAACCGCGCGGTGTTTCCCGCATAGGCGTTGCGGTACAGGGTT
262214 |----| |----| |----| |----|

256716 TGTACCAACGATCGCGCCACCTTCGCCGCCCTTCGGCGAGGCGTGTCCGATGGACAA
|-----|
256776 CCTGATGCGCGCGGAGAGACGCTCCGTCGCTTAAGAGAGCGCAGGCTTTGCCGAGCCT
256795
256836 TTAGATTTACAGGTAG
...
256863 AGCATTTCTCCATGCGCGCGGCCCTTCGGGCCCTTCGTAGCGGATGATGACGATCG
|-----|
256913
256919
-| |-----|
256923 CCGCGAGGATTTGGTTGCCCAAAATGCCTTCTACTCGCTCTCTTGGCTTTCAAACACG
256925
256924 256933
256983 C
...
257001 TTGAGGATGCTCTGTCTCCAGCGCTGCGGTTTTTACCACGAGCGCGCTCGCGATGTG
|-----|
257051
257061 CCGAACAAGACCGCCAAACCGCGCTCTTGCAGTAGCGGTGTGCCA
...
257119 AGCCTTTTTCGCGTCTCGAGGTCGAGGTTTTTCCACATACGTTTTTTCGAGACGCTTTGG
|-----|
257169
257179 TGTGCTACGCCCGCGCGCGCTTTGAAGCGTTTCGATGCGGTGGTGTTCAGGGT
|-----|
257209
257239 TGTTACGTCCCATTTTCAATCGCGTCTTTCACGCTCGCGCGGTGATAGTGTGCACGT
|-----|
257271
257299 CGGTGTCAGCTTGCCCGCTTTGTCCAA
...
257487 AAATGCAAAATAGTTGTGTGCTGCCGCCATGSCAATGTCCATGTTCCATAGCTTTTCA
|-----|
257537
257547 AACGCTTTTGTGGCAATGCTCGCGGTAACACGTTTTCATCGTTTTTGTCTCGTAATAG
|-----|
257588
257607 CGTTTGGTGATTTTCGACAATCATACGCCGCTTCGAGAGAACAGTTCTTTTCGG
|-----|
257610
...
257727 TTCATCGAGTTTGCGTAACATACCGAACACGAGCCGAGGTCGGGCAGGCTTTTGT
|-----|
257777
257787 TCGACTTCTTCAACTTCCCGGTTGCTGATATGTCTCGTCCGCGGATTCAATCATCGCGTCA
|-----|
257819
|-----|

257847 ATCAAGTCCAAACGGCGTTCCGGCTGGATGTTTCCACGCGGATAAATTTACCTGCTTCC
257860
257907 ATCGGACCG
...
258048 GCAATCAATCGCGCTGGCGAGGAATACAGCATACCGCTGTGTCCCATAGCGATGCCG
|-----|
258098 258107
258108 TCCTCGATGCGGATGGTGTTCAAATCTTTGGCGATGGCGCGGCTTTTTCGATTTCCGGG
|-----|
258112 258121 258137 258143
258168 GCAACACGCTGGCCCATATTGTGCAGGTGGACATGACCTGGCAGCAATTTGGGTGAACGAG
|-----|
258228 TTGGCAACGGCGATAATCGGTTTCCGGAAGTCGGTTTCCATCAGCGCGGTGGCGGCCAC
258236
258288 AAG
...
258541 ACGGCTTTGGATCGTCCCAAAATCAACGGGTCCATTTAATACGGGTTTTCACGATGCC
|-----|
258591
258601 GTCCGATGCGGAAATCGCGGAAACCTTTCGCCGATATTGGTCTCTCGCGCCATTTCGCG
|-----|
258615
258661 ATCAGGTCTGTCCAATTCGGAAGGATTTTCGGGAGGCTGATGTTTTCTTTGTAAGACGG
|-----|
258666
258721 GGGATGCGTACGCGCTTTCACGCTCGCGCCGATATGGAGTTGTAGCGTCCGACGCTTTG
|-----|
258741 258764
258781 CCGACCACTCGGATTTCCGCCAACATCGCGCTCCGACGCGTTCGGGACGCGGTAATG
|-----|
258805
258841 CGGTAACGATGTAGTCTCGGACGTCCGCTGTTTCGCCATAAATCTTATCCAGCTCGCG
|-----|
258846 258855
258901 ATGAAGTCCG
...
259051 CCTTCGGGTACATTTGCCACGATGAGTTTTTGGTTGGCGGTGATGCGAGTCCGCTTTG
|-----|
259101
259111 TGGATTTTGGGATTTCCAAACACGCGGTCAGAGCTGTTTCCGCGCTTCGTCAACCAAA
|-----|
259119
259171 CGCCGCTTTTCGATGAAGGTCAAATGCCAGTTCGCGCTATGCGCTTTTACCACCGCG
|-----|
259174
259231 ATGGGTCGCGCGCGGTAATTTTGAACGGCGTATGGTTCGAAACGCGATACCCATA
|-----|
259236
259291 CGCGTTTCGACTTCCGCGCGGAAGTTGTCCAAGCCCATATTTTGAATGGTGTAGCGGGT
|-----|

```
254051 CGCGCTGTAAATACGAAGAACACACATTATGACCGTATCCCCGTA
...
254317 CGCACCAACACACCAATCTGGGCAAGAGCAGATGATGGATGAAATCGTCGAAA
|-----|
254367
254377 AAGCCAAATCGCCCAATCGCGCGCAAGCCGTTTGTATGGCGCGCATGGCGG
|-----|
254385
254437 GCCC
...
254560 CGGCTTTGGATTATTACAACCAACACCTCGACCCGACCCCGCTACACGACATCA
|-----|
254610
254620 TCCATACACCGCGCATGAACCGTATGACACCTTGGGCAAAAGTCCGCAACGCCGTT
|-----|
254670
254680 TGAAGTCTGTGCGGCGCATCGTCGGGATGAACGAAACTCGTGCCGAACGCCCGGC
|-----|
254700
254740 TGATTGCCAGCCTCGC
...
254773 CCGAAACGTGCCGATTAAACCGTTGGTCAAAGTGAAGCCACGCCGCTTGACGATGCCG
|-----|
254823
254833 AAGATTGGACTGGACGAAATTTGTCGCACCATCGCCGTGGCGGCAATTACGATGCCG
|-----|
254865
254893 AAAATTATGTCCGGCTGTGCGAGGCGCAGCAATATGCCGGAAGGATGACGGCGATGT
|-----| |-----|
254936 254945
254953 GCTTTATGGCGGGCGAATTCGATTTTACGGCGACAAAGCTGTGACCACAGGCAATC
|-----|
254985
255013 CGGACGAAGACGGCGACAGAAATCCTGTATGAAAGCTTAACCTGTATCCCTTACAGTTG
|-----|
255024
255073 AGCTGGAAGCGAGGTGCGCGAAGTGGAAGCCCTCGGGATTAAAGTGGATTATTGAC
|-----| |--
255087 255130
255133 GATTGCAAAATGCCGTGTGAACCCCGAAAGGCTTTCAGACGGCATTTGTT
--|
255135
...
255216 CGCGTGATGTGCCGTGGGCGAGACATCCAGCCCGTTCCTCGTTGGGAAATCGCTGCGG
|-----|
255266
255276 TTCAGATAAATAATCCGCCGATAACGGTTTCTCGGGTGCATTTTGGGGATTTTATCG
|-----|
255332
255336 TCGGCTTGGGTAATAGGATGATAGTTCCGGAAATAAATTTACCCGTGTTGT
-|
255337
```

```
...
25525 TGCCCCGAATACGACATCCCGCGCGGAATCGACGGCAAGCGCGCAAACTCAACGCGTTG
|-----| |-----|
255575 255578
255585 CCCCTGTTTCCCATAGAAAGATTTCGGTATCGAACGGCAGGTGTTCCGCGCTGTTCCCT
|-----|
255645 TCTCGCGCGCGGTGGAACATCACAATAACGATCTCGGCTTTTGTGTTTGGCAATC
|-----|
255671
255705 AGTTTTTTGACTTTGGCATAGTCGTTACGTTTACGGCGGCAAGGTTGGGGCGAAGGAG
|-----|
255725
255765 ACGAAGCGGATCTTA
...
255927 TCCCCGTGTGCGGATTGACGGTTCAAACCCCGCGCGCTTTCAAATACCGCGGATGTGG
|-----|
255977
255987 ACGACAGGATGCGCCTGACGAAAGACAGCCGATACCGTTTCCATTATCGGCGTGGCG
|-----|
256043
256047 ACATTATGCTCGCAGCAATTATCCGTCGATTACCTGCCGATACCAATATTCTGAANA
-|
256048
256107 ACGTCGAATCTGCTTTCGAAGACGGGACATTACCCTCGGCAACCTCGAAGGCAC
|-----|
256111
...
256176 GCGGTACGCCGAAAAAATGTGCAACCCCAAAATATGCTATGCATTCCGAAAGCCCTCC
|-----|
256226
256236 GCATACGGGCAATACCTTGCCGACGCGGGATTTCGACTACCTCAGCTTCGCCCAACACCAC
|-----|
256299
256296 AGCAACGACTTTCGGCGCGCAAGGCATCACGGCAACGGCGGAGCGCAGCTCTTTTAC
|-----|
256415
...
256365 GCGCTAAAGCCGTGCCGATACGAGGCCAAAAATTCCGGAAAAATACCGCCATCGCCGAGA
|-----|
256415
256425 TAAATTTGCCATCATCAGACCTTTACTGTTTCAGACGAGACAGCAT
...
256489 CTTATGTTTCGATTTCGCTAGTCGCGCACCGCGCTTTGTCGGCGAAGTCGCCATCG
|-----|
256539
256549 CGCCGTAAGCTCTTAATGCGCGGAGACCTAGCGTFCGCGGTTTC
...
256656 TCGAAATGACAGGCGGATGCTGCGGTTGGGATGTCGATTTCGACGATCGCCTCG
|-----|
256706
```


251317 AAAATCTATTACGGCGTATTGATTTCTTATGTATCGCTTCTATGCTGCTGCGCGTTT |-----| |-----|
251350 251367
251377 TTTTATCGGGTGCTTTGAAGCCCAAGAAGCGGCATTGCGGRAAG
...
251569 AATCAACGAATTAAAGTTTTTGTGAAAACAAACACAGCTCTAAGCAAAATAGGCGGTTTGTG |-----| |-----|
251619
251629 GGTAAATACGGAAGAGTTGCGGCATTATCGGGCATCTTTAACAAAGT
...
251708 GCAGCTTGCAAAATCGCGGATATAAAATTGCGGAATCGGTTAAAGTGTGGGACGCTATGA |-----| |-----|
251758
251768 AAAATTGCGAATTTTTTTATGACCCGACAAAGGGCAATCTATGATAG
...
251906 ACGAGGTTTTTGGAAAAGCGGAAAATACCGTTGCGCAACACAGGAAATTGATGACGATATGG |-----| |-----|
251956
251966 ACAAAAGTGTCCGCCCTTAAGCGGAAGTTTAAAGATATTTCTGAAT
...
252060 GAAAGAATTAGATGAAGTGTCTATTGAAATACATTCCCTTTTTTTGAAGAAGAAGATAAAG |-----| |-----|
252110
252120 GAGTTGATATGCCGCTATCTAAATAATTGGAAGTATGTGCTTG
...
252289 ACAGTTCGGTTAGACCGGCAGTAAACACACAGGGCGGAAGACGAGGGGTGCGGATTAAT |-----| |-----|
252339
252349 CGACACGGTAGGGACGACACGCTCCAAAATCCGAGAGAGATAC
...
252407 TTCGAATACCGGTGCAAAAGACGCATTGATTGAAACCGAATCAGCCATGACGGCGTTGAAT |-----| |-----|
252457
252467 AATAGAATTTGACGATTAGGGAACCTGTGATGAACACATCCACATTATCGGTATCGGGC |-----| |-----|
252478
252527 GCACGTT
...
252638 TAGCGGTGATGAAGGCTTCACACCGCGCAGTTGGACGAATTTAAAGCCGACGTTTACG |-----| |-----|
252688
252698 TTATCGGCAATGTCGCCAAGCGCGGATGGATGTGGTTGAAGCGATTTTGAACCGTGGGC |-----| |-----|
252738
252758 TGCCTTATATTCGGGCCCGCAATGGCTGGCTGAAAACGTCTGCACCATCATTTGGGTAC
...
252818 TCGCGTGGGGGACGACGCGCAAAAGACACCGCGTCTATGCTCGGTGGGTTTTGG |-----| |-----|
252843

252878 AATATGCCGACTCGCACCGGGCTTCCTTATCGCGCGTACCGGAAATTCAGCGTTT |-----| |-----|
252931
252938 CCGCCCGCTGCGCCAAACGCGCGCCGAAGACCCGAACAGCCGATCGCGGTTTTCGTCA |-----| |-----|
252954
252998 TTGAAGCCGACGAATACGACACCGCGTTTTTCGACAAACGCTCCAAATTCGTGATTAAC |-----| |-----|
253034
253058 GTCCGCTACCGCGTGTGTGAACAATCTGGAA
...
253127 TACAGACCCAGTTCCACCACCTCGTTCGTACCGTACCGTGCCTGTGAAGGCTCATCTGCTGCA |-----| |-----|
253177
253187 ACGACGGCAGCAAGCCTGCAAGACACTTTGGACAAAGGCTGTGACGCGCGTGGAA |-----| |-----|
253233
253247 AATTGCGCAGGAACACGGCTGGCAGCGCGGCGAAGCCAATGCGGATGGCTCGTTGACG
...
253307 TGTTCGTTGACGGCAAAAAGCCGACACGTCGCTTGAGTTTGTATGGCGGACACAAAC |-----| |-----|
253336
253367 GCATGAACGGCTCGCGCTCATCGCCGCGCGCTCATCGCGAGTCGACATTCAGACGG |-----| |-----|
253387
253427 CCTCGAAGCCTTGAGCACGTTTAAACGTCAAACGCGCGCATGAAATCAAGGCACGG |-----| |-----|
253453 253460
253487 CAAACGGTATCACCGTTTACGACGACTTCGCCACCATCCGACCGCTATCGAAACCAGCA |-----| |-----|
253507
253547 TTCAAGGTTTTCGCCCA
...
253652 GCCTCAAAGAAGCCGACCAAGTGTCTGTACGCCGGGCGGAGCTGGACGTTGCGG |-----| |-----|
253702
253712 AAGCCCTCGCGCTTTGGGGCAGGCTGCACTGCGGCAAAAGACTTCGATGCCTTGCTTG
...
253772 CCGAAATCGTGAAAACGCCGGAAGCAGCGGCATATTTTGGTGATGAGCAACGCGGTT |-----| |-----|
253786 253798
253832 TCGCGGGAATACACACCAACTGCTGGACGCTTTGAGATAGCCCGGCGATGCGCTCTGA |-----| |-----|
253858
253892 AAGCCCTTCAGACGCGCATCGCCGCGCTGCGGGCACAAGGCGGAAACCGTTTGCCCC |-----| |-----|
253908
253952 GTATTTTCAAC
...
253991 GCCATTTGATTAACCGCAACCGCGCTCATTCGCGCGGGGTTCTGCACGTCAGCGCTATTG |-----| |-----|
254041

247245

247271 CGCGAAACCATCGATCCAGAACGTCGCGCTTTTACCCTGCGGACCTGATTATGGC

|-----|

247331 ATTGCCGAATTTGACGACGACCCGCCCGTTGCGGGAACGATCCTCGAAGCGTGCA
247344
247347

|-----|

247391 GCAGGCATGGAT

...

247609 TGCGTTTTTCGCGTCCTTCATACAGAACGCGGCTCGGATCCCGGGCGGTTCGGT
247659

|-----|

247669 GTTCAAACTTTAACTTTGATTTATAGGAAGGAATTGAGGCTCAGGTCGATAATATC
247709

|-----|

247729 GCGATGTCATGATGTTTACTGTTTTTGACCTTCGAGCCGTTTACTTTGAACCC
247731

--|

...

248181 AGTCGCCGTCGATATTTTCGACAGAAATACACCGTTTGCAGAAAAAAGCAACGCTGA
248231

|-----|

248241 CCAAAATCGATTTTCAGCAAACTCACACCCGCCCAATTTACAGGTTTCCCGGCATCCG

248301 AGGTCCTCTATACTTTGGATTACATTGAGCACTGTTTACCGATTTTGAAGAACTGCACG
248301

|-----|

248361 GCGACGCCCACTTTGCCGACGATTATGCGGATGTCGCGGATTTGGCGGTTTCAACGGAC
248365

|-----|

248421 AAAGCGTGATGGTCGTCGCGCATCAGAAAGGCGCGACACCAAGAAAAAATCCGCGCA
248432

|-----|

248481 ACTTCGG

...

248670 CTGTTTGTGTACCGTTATCGCGAAGCGGTTTCAGCGGTGCGTTGGCGGTGCGCGTAG
248720

|-----|

248730 GCGATTACGTCATATGCTGCAATCTCGACCTATTCTGTATTATCTCCCCGAAG
248734

|-----|

...

249178 GTCTGAATGAAAAAACAAGCGGTGGCAATTAAGCGCGGCTTGGATTCCGTCGTTTGC
249228

|-----|

249238 TGCATCTGCTTGTCCGCGCGGAAAAAAGGCGGTTTTTATTCGGATGCAATTCATATCC

249298 ATCACGCTTGAGTCCCGTGCAGGATTTGGCGAGATTTCTGCCAAAACTATTGCGATA
249321

|-----|

249358 TGCTCGGGTGGGCTGGA

...

249386 GGTCTGCTGGAAAAAACGTTTGGCATCGAGCGCGGCAAGGCAAGCGCTTATGC
249436

|-----|

249446 CGCGTTTGGCGAAAAAGGCTTTGACGTTTGGCGTTGGCGCACACAGGACGATCAAT
249468

|-----|

|-----|

249506 CGAAACCTTTATGCTGCGGTCGCGCGCGGCGGCGGTTTGGCGCG

...

249772 TGCTGAACAATGTCCCGGCTTTGCAGGAAGATTTGGCTTTTGGACGAGGTCTGCTTC
249822

|-----|

249832 AGGACTGCCGTTGGGTTTGGCGGCCGGTTATTTTCGATACGGCGGTTGGCTGACGTTT
249884

|-----|

249892 CCGCGCGCGAAAAACCATATTTTTCGCGCATTTTCTGAAGAAAAACG

...

249956 TCAGAAATGCCTTTGCCGACATGCCCCGGTTTTCACGGAGGCAAAACCGAGCTTGAA
250006

|-----|

250016 CTTGCAAGGCTTTGAATTGCATCATTTATGCAGCAGGCTGTTTGTG

...

250299 GTCTGCGCCCAATTTGGCGCTTGTTCGACAGCGAAACCGTCCATTGGCGTTGGCA
250349

|-----|

250359 AACTGTTGTCGGGATTTCCAATACTCGGTTTCAGACGGCATTTTGC

...

250410 CATCTGACCTTCCCATTTTATTTGATAATATCGCAACAGATTTTCGGCGGCTTCAGT
250460

|-----|

250470 CGGATATTGTCGGTTGCATATTTCTAAAAGGCTTTGTAAAGTGA

...

250793 ATGTGCTTGGCTGATGATGATATTCAGATCGGGTGCTTTGCCGGCGGCGGATTCG
250843

|-----|

250853 TGGTTTTGGACGGCTCCAAGACCCGGCAATGTCGCGCACGGTCTT

...

250912 GCGCGGGAATCGCGCGGTCAATTTTGGGCAAGTTTTCGGACGGTGGTCGCCCA
250962

|-----|

250972 GTGCTGCGAGCGGAATTTGGCGCGCATTTCTTTCGGAGATTTATC

...

251197 GTCAGGATACCGATGCACGATGCAACGAGCTTTTAAATGTCGGATGGCGGACATC
251247

|-----|

251257 TGCTGTTTGAACAAATTCGCCAACGGCGGCTATTGAGGAAGAGAAATGCGCTCTGAA

243900 243900
243907 ATTTGCAAGCCGCTCTGAAAGACGGAAGCCGAAGACATCCGCTCCGCGTCCGCAAAAC |-----|
243957
243967 TCAGCCGCAGACGCACTTCGCCGCCAAACGATGAACCGCAA
...
244160 TGAACCGTCATCAAGCGGTGTATAGTGAATTAATTTAAACCGGTACAGCGTTGGCT |-----|
244210
244220 CGCCTTAGCTCAAAGAGACGATTCCTCTAAGGTGCTGAAGCACCACCAAGTGAATCGGTTCCG |-----|
244237
244280 TACTATTTGTACT
...
244358 TAAGCAATTTTGAAGGCAGAAATCGCCATTTTGGAAACAACGGCTATGCGAAGCTTGCCG |-----|
244408
244418 GATTGGAAACGATCCTCGCGCTCAAAACCGACAATTTACAAGAAAAAAGAAACCTGTCTC |-----|
244478 ATGAAATTTTCAGCAAAAGCAGGATAGCGCATACCGGAATTTGTCGCCGTGGACCAAAAC |-----|
244504
244538 TTGTAAACGTTTGTGTTTATCGCTTCGCGCGGAAATCGTGTTCGGAAGAAACGAAACGG |-----|
244544
244598 CGGAATCGGAATTTGAAAAAGCGTCCACGAACGCGAGGAGGAAGGAATGCGCGGTCCG |-----|
244618
244658 GCGCAAAACAACGCT
...
244934 CGCAGTAAGGTGCAAAACAGATACCGTACAGTTTCAAGGACAGATATGATGGCGGATTCG |-----|
244984
244994 TCATTTATTATGTAGAACGCGAGACAAACCGGTAAACCGTGCCGG
245040 CGAACGCGCCGCAAAACATTTCAATATTTTGGCGGAGCTTTTGGGAAGCGCCCGG |-----|
245090
245100 CATCATTCGCGCTTAGATTTGCCATGGTCAAGCCCTTTTTCNAAGACGGCGGAAGC
245160 GGCAGAAATTTGCGAACATATGTGGATCGACGATATTTTGTACGCTCTTTACATTTAT |-----|
245187
245220 GCGTGTGAACAAATGAACCCGGGAACTGACCAATGTCAAGCAAGCGGAAGCGGTTGC |-----|
245271
245280 GTTCCGTTGACGACATCAGCGACTGGATGTTGTTGCAACGGCATTCCTACGGCGG |-----|
245289 245298
245340 TTATCATACAGC
... |-----|

245368 TGACGGAAGAGGACGCTACCGAACACACGATCGCGCATGGGAATCGATTTCCGGGATCCCG |-----|
245418
245428 GGCAGGTATTGCTGTATGAAGAAAAAGAACATCCGAAAAATCTGGAAGAGCATCCGA
245488 TGTGCCGGAATGTATTTGACGATTTTCGGCAACAGTTGTCCCAAAATCCGATTTCTGC |-----|
245505
245548 ATGAGCAGGACGA
...
246016 ATCAGGCTTGGGTTTGGAAAGCCGATTTGCGCCCTGAGTTGTAGCGCGGTTGTCCCGCGCG |-----|
246066
246076 AAGATTTGATTGTGGAATCCCAATACACCATCAACACGCGCG
...
246143 CGTCTGAAGCGCGCAGCGCTTCAGAGGCATTTGTCGCGGATAGGCGCAATCCGCCGAA |-----|
246193
246203 AACAGCGCTTCGTACAGCGGAACTTTTCGATTTCTATAGCAATTTAACAAAAATCAGGACA |-----|
246207
246263 AGCGCGGAGCGCGAGACAGTACAAATAGTACGGCAAGCGGAGCGGTGTACCGTT |-----|
246308
246323 TAAATTTAATTAATTAATTTATGATTTTATCGGTTTTCG
...
246442 AAACCGCGCGGTTTACTTTACGGCGCGGTATGCGGCGGTATGCGGATAGCGTCTCCAC |-----|
246492
246502 TTTCCGCTCTATATTGATTTTGTGTTTCTGACGGAATGACCC
...
246604 TCCTTGCGTGAACCGGAAGAGACGCGCGCGGACAAAGCTCGAGATAGCGTCTCCTC |-----|
246654
246664 AAGTCCGGCAGCGCGGACACGCTTTCTGCGTTAACCGTGAAGCGCGCGCCGACCC
246724 GGAAGCGGTTCCGCAATCTGCGTGTAGATACGCCAACGTTGGCGCGCAACATTTGCCG |-----|
246729
246784 ATCGCAACCGCGACGCGCTTTATCCAAAGCA
...
247042 GGGCGGTAGAATCAGATTTGTTGGAGGGACAATCTCTCCAGATCAGAGCAGACAT |-----|
247092
247102 AGCTGTGTGTTGATGTTGTCCGCGAGTTGAAACATTTCAGCA
...
247151 TCAAGGGCGGCGAGTTTTCGCGAAACATATTCTACAGGCTTCAATGCCGCGAGATAAAA |-----|
247201
247211 GGAAATTCATATGAATGGACCGACACCCAGCGCATCGCGCAAGAACTCTATGACCTGCA |-----|

240981 TGCCGGCCCAAGGTAACGGCGGGGCAAGCCGGCCAGCCTGGCGGATGACGGGGATT
|-----|
241034

241041 TTCGGGGCGGATCGGGCGTGTTCATCATTTATTCTCCAAATGGTTGA
|-----|
241237

241187 CCGAATAACACAGGTTTTCAGGGCAATCCGTTTTTTTGGCGGCGAAGGAGGGGATTTTG
|-----|
241237

241247 TATTTGAGGACGTTTGTGCGCGTCCGCGTCGATTTCGTTCATAGCTCGTAATAAAC
|-----|
241255

241307 GCCGCCATAACACAGTCCGACTTTTGGGCTTTTTATCGGCAGCCGGCAGCAGCGATACG
|-----|
241358

241367 GCTTCGGGTAGGTTTCACGGGCGGTTTGATTGGAACGCCCATCAATTCCGGCAAAATG
|-----|
241403

241427 CCGTCGGGCTGCATTGCAAAATCAGGCTTGC
|-----|
241403

241533 GGATGATGTTGGTCAGTTGACGGCAAGTCCCATCTTGTGGCGTATTCACAGCGTTTGGT
|-----|
241583 241591

241593 CGTCTGAABACCCCAAAATCCGGCAATCAGGCAGCCGACCGCCCGGACGC
|-----|
241596

241775 GCCTGATTACGGGGTGTCCGCGATCGCGCCCGCAATACCTTGTCCAAATCGCGCGC
|-----|
241825

241835 CACCAAGTTCAATGTCGCGCTGTCAACATCGGGTTGGAACATTCGTCAACCAATCGTCC
|-----|
241847 241888

241895 AATTCCGGCAAAAGCATATAAAACCGTTACCGCATCCGTTTTTCCT
|-----|
241888

242308 GCAGTCGGCGCTGCCTGCCGATGCCGAAGGGCGGTTACCCGCTTCCGTCGTCGCT
|-----|
242358

242368 ATCTGAAAACGGCGGCAATTGAAGTCGGCAAAACCGCCTGTCCGCCCAAAACCGGAC
|-----|
242428

242428 CGCTGAACACCGTCAGCTCGGCCCAACGCGTTATCGGGGAGCGCTTGGGATCTGCATC
|-----|
242452 242468

242488 AAAATACGCACTACCTGCCTTACCGTTTCGGGACAAATCAACGCGTTATCGAATGCATA
|-----|
242517

242548 CGCGGAGGGGTGAACAGCGCTGTGCAAGTGTGCGCGGAAATCCTCAAAACCCCTTAAT
|-----|
242564

242608 CGCGGCCGGAAGAACCTTGGGCGGCGATTTGGTCGGCGTGGTATTACCGTCCCGCGCT
|-----|
242631

242668 ATTTTCAGCATGCCCAACGCCAGCCACCAAGATGCCGCGTCTGTGGCGGTTTGAACG
242673 242683

242728 TATTTGGCGCTGCTCAACGAACCCACCGCCCGCAATCGCCTACGGGCTGGACAACGCTT
|-----|
242763 242781

242788 CGGAAGGCACGTTTGTGCTAGGACTTAGGGGGCGGCACATTTCGAGCT
|-----|
242904 242907

242914 TCGACCACCGCCTGTTCTGCCGCTGCTGCAACAAACGGAAGCTCTCCCA
|-----|
243000

243060 TCGAGCGGATGCCAATCGACACAAGCATCAGTCGCGCGAGTTCCACAACCTGACGCAG
|-----|
243120 243134 243155

243180 AAAAAAGAGTCAAGGCGTGATTATGTTGGCGGTTTCGACCCGTTGCTGACGCTCCAA
|-----|
243240 243246

243300 GTCGCGCTCGGCGCCCATACAGGCAACGCTCCTCGCAGGCAACAAACCGGCGGAA
|-----|
243302 243309 243327

243360 TGGCTCTGCTGAGCGTTACACCTTTGTCGCTCGGTTTGGAAACCTACGGCGGCTTGGCG
|-----|
243372 243386

243420 GAAAAAATCATCCCGCGCAATT
|-----|
243453

243513 GTCGTCAAGGCGGAGCGGAGCTCGTTTCTGACTGCCGAGCCTTG
|-----|
243667

243727 CGATTACCGGAATGCTTAAAGACAGTATGGACAACGCCGCCGAAGATATGGCGGCACGG
|-----|
243759

243787 CCGCGCGCGAAGCGGTAGTCGAAGCGCAAGCGTGACCGATGCCGTCAGCGCGCCCTCG
|-----|
243834

243847 AGTTGGACAGCATTTTACTGGATGCCAAAGAAATTGCAACAGATCCGGCAAGCATCGCG
|-----|
243894

237730 TCAAGGTTTCATCGCTTGCCTCCTTTACCGGTTTTTCATCAGGAAGTCCATAAATACGCC
|-----|
237741

237790 CGATTTCG
|-----|
237917 GCATTCCACAAAGTATTTTCCATCTTTTTCGGTCTGCTTGCCCAACGAAGTCTGCTCGTTT
|-----|
237967

237970
|-----|
237977 TTCAAACTTCATAGACGGTATCGGGATAGCCGCCGTAATAATACAGGG
|-----|
238026 ATTTTTCGGGTTGCCCGCAGCGGTGACAGCCAAGACCGCGCACAGAGGACAAAC
|-----|
238076

238086 GGCTCAAGGTTTTCGGATTTCATCTTCTCTTAACGGTTGGGTTG
|-----|
238244 ATGATTTACGGTTGGAAGTGGGTATTCGCCCGCCCTGTGCGGAATAGACGATTTCG
|-----|
238294

238304 GAAGTATTGACGTTGACGATATTACAGACCACCTTTTCATAGCGGATTTGCGG
|-----|
238312
|-----|
238318

238364 CGACCCAAAATGCCAAGAGCTGATGATCGCCGACATCTCTGCTCGGAATTCGGTTACA
|-----|
238390

238424 TCGCCGTACGACATAAATCTGCGCTTTCAGGTTATGCGCTTTGCGGAAATGCCGGAT
|-----|
238428
|-----|
238437

238484 TCCT
|-----|
238637 GTTCCGACGGAATCGGGGTACGAACACCGTGATATTGCGTATTTAGGAGCGGACTTTC
|-----|
238687

238697 TCTACCTCGAGACTGCGTGAGGATTCGGTCGCACAGCCGGTCAGTG
|-----|
238912 ACGGAATCGGGGTGGCGGGATGCTGCTTTTCGGATTTAGAGGTAACGATTTCG
|-----|
239132

238972 CCGCTCCGCCCTTTGCTTTTCGGTACTTGCCACACGACAAATGCGG
|-----|
239109

239059 CCCGGATGGATTGTATGCGTTTCGGGGAATTGATGGTCCGGGACGAGGACGTTGGCG
|-----|
239109

239119 CGCAGGTTGCCGAAGGTTCCCAATTCGTGCGCGGCGACTTTGCACAGGTAGTTCAACGCG
|-----|
239132

239179 GCTTTGGACGCGCGAAGCGGCCAGT
|-----|
240971

239210 CTTTGGGTGTTTCGCCGTGGCTTTTCGCCGACGAAGATGACGGACCGCTCGCGGCGACTGCT
|-----|
239260

239270 TCAGCAGCGGGAACAGCGCGGGTCAAGCCCATAGGTGCGACGGTGTGATCGGGTATT
|-----|
239330 GGTGACCCATTCGGCGACGCTTTTGGAATCCAGCGCGAGAGGGCGCTAAAAATAGCCGG
|-----|
239343

239390 CGCAGTCGAGATGCCCTCCAGTTTGCCTTGGCTTGGGTGCGCAATGGTGGCGGCGAACT
|-----|
239397

239450 GAT
|-----|
239554 TGATGGCGTCCACACAGAAATCAGGTTGCGCTGCCCGCATAGGCTTTGGCGACCTGT
|-----|
239604

239614 TCGCCCGACCTTTCGATGCGCGGTAATTAAGATGGTTTGTGCGGACAGTGTGCGCCATA
|-----|
239665

239674 CTTTTCCTTTTGGTGGTGGGGTATTTTACGCTTTTCCCGCACCTTGTAAAGC
|-----|
239709

239734 GTCCCGATACCGCGCGCTTTTCAGACGCGCATCGCTCAGGCTTTGACGAGGCTTCCG
|-----|
239736
|-----|
239749

239794 CCAGCGGAACCTGATTGTACGGCGCTTCGAGCGTGGCGGGGTAGTCTGGGTGGAGGT
|-----|
239855

239854 AGTCGCGCGGGGGAAGATCGGTTGCCGTTGCAACACGACAAATCCGCGCGGGGG
|-----|
240000 TGAATTTGTCCGCCACGCCGCTTTGCAACGCCGCGGCTCGGAAAGCTGATG
|-----|
240050

240060 ACGGCGGACACTTCGTTTTCAGGCACTCCGAGCTGCCCGGCAAA
|-----|
240363 TCAGAGGGATGTCCGGCGGAGCGCTTGAAGCTCCCGCAAGGGGGTTCCGGGACGATT
|-----|
240413
|-----|
240416

240423 GCGCCCAATCCCTGTTTGGTAGAGATAGTCCGCTGCCGGATTTTTCGTACGACGCCG
|-----|
240452

240483 TCAGACAAACGTTGCACAACACGCGCAGGCTTGGCGTTTCCAAAGCGGTGTGAGCGCG
|-----|
240491

240543 CCCC
|-----|
240921 GCCAGTGGCGCGCCCTGCCGCCCGCTTCCAGAGGGGTAAACAGGGTAACGTGCGCG

234345

234386 CCGCGCCGAAGAATA

...

234646 ACAGCTTTTACAGGCAGGAAAAAGAGCCCTCGCGCGGAGGAGGAGCAACAACGCGCCGA
|-----|
234696234706 CCCCAGTGTGGCGCAAAATCGCCCGCGCAAGGGGGGTGGAATTGGACGTTTTTGAATGAAA
|-----|
234723

234766 AAGTTGTGCAAAAATCCGCCCGCCTGGCCGTTGCGCGCCGG

...

234852 ATCGAAACCGCCAGGATTGGACGGCTGGAAGGAATCGAAGAATGGACGCTAAAC
|-----|
234902234912 ATCGGCTGAAAAATACGTTTACCACCTGTTGGTAGCCATCGACCAACTGTTTCAACGCC
|-----|
234965

234972 TAACCGCGCGCGGGAGCAAAACCCCTCTCAAGCGGCACCTATCGCGG

...

235278 ACTGCCCTACTTATCTTGACCTCGTGCAGAACTGGCCCTGACCGGATGGGAGCGTTTGC
|-----|
235328235338 AGGCAGCCAAACCTTGACTGTGAGCAGCACATCAACAGCAATTCAGCTACGACGACCCAC
|-----|
235389

235398 CTGATTATAAGTTCTGACCGCAAGTAGCGTACTACTTTTAAAGCA

...

235607 GGTTGCCGCGCGCTGTGTGTTCTTCGTACTCAGCAAAAAGCGGTACAGACGCTCCAA
|-----|
235657

235667 TCGGGCGGGTAGCGGGTGAATTTGTTAGAACATCCGGAAAAAGATAGCGCGTTTG

235727 CAGTCGCGCAAGGCTTGGACGGTCTGCTGGATGTCGCCGATTTTGATTGTGCCGCAAA
|-----|
235760235787 CAGGCGCGCGCTTGCAAAAATAATGGGAAGATTGTGATGCCGTTGGTGAACATATCGTT
|-----|
235794235847 AAGCGGCTTAAGCAGACGCTTTGGCGCGGATGCCCAACGGTTGCGGATAATGGCTTT
|-----|
235862235907 AAGCGGTGGAAGACCGCGCGTGTCTGTGTTGTCGCCGTGTGTAACAGCTACGCTTTC
|-----|
235953235967 GCGTGGTTCGCGTACGCGGATGAGGAATAACGGTAGTCGCGGTTGAGTTTTTCGTTTTC
|-----|
236003236027 ATAGTTGTAACGAATCAAGGGTTGCCCTATCCACATCGCGGATAAAGGTGCGCAAAATCAC
|-----|
236073

236087 AAAAAAAGACGAGCCAAACAATGCCGCGGGAATGTCGAA

...

236139 GTCAGATGCCAGCCCAAGCCCCACAAAAACAACGGCAAAATTTCCAAGAGAGTAACGACCGAA
|-----|
236189236199 TTGACCATGCCCGGTACAAAATTCGATGGTCTGAAGCGATGAATCTCTGCCGCTCCTGTTGG
|-----|
236231

236259 ATACGTTGGTCGATGTTGTGCGCGCGTGGCGCGCATTTGCAGGCGGTAGTAGTTTTTG

236319 TCGGCAAGCCAGCGCGTGGTCAGACACTTCATTGAGCTGCTCCGACCATTTTAATCCGCAAG
|-----|
236334236379 CCTTGATCGAGGAAGTCTGTTGACGACATTTGTTAAACGCCCGTATCAGCACCACGCCCGG
|-----|
236393
236399
236412236439 TTCATCGCGGCAAAACATCCAAAAACGCGGAGGGTTTCAAAATCCTGCATCGAGTCGTAAGT
|-----|
236468

236499 CCTTTGGACATAAAGGTACTCAACA

...

236736 GTCAGGGTACATGGCGTGTAAATTTATAGTGGATTAAACAAAACAGTACAGCGTTGCCCT
|-----|
236786236796 CGCCTTAGCTCAAGAGAACGATTCTCTAAGTGTCTGAAGCACCAGTAGTCGGTTCCG
|-----|
236813236856 TACTATTTTGTACTGTCTCGCGCTTCGTCGCCCTGTCTCTGATTTTGTAAATCCACTATAC
|-----|
236881

236916 CATACAACCAACGCCGAATTA

...

237103 GCTGCTCTTCGACGAATCTCGGACCTTTCAAGATGCCGTTGTGAGAATAGGCGCAGCA
|-----|
237153

237163 GGTTCGTATGGCGCGGTTTTTAGAAACCTGATAACCGCGGTTCGTCAG

...

237307 ACAACTCTTTTCCGTTGCGGNAATCGACCACCGCTGCTTTGCGCGATAGCGTCTCACGC
|-----|
237357

237367 TGCTAAAATTTGATATGAAGTGGCGTATTCGGTAACCGTAATGTA

...

237550 AAAGCGCGCGCGGTCCGAAGCCAGTACACCCCATGTTCCGTTGACATCGGCGGATCGT
|-----|
237600

237610 TCAGCGCGGAACCAACCAAAATTTGAAGCGGTTTGCCTTTCCCTTGAATGACGTTAGTCGA

237670 AATCGGGCGCTTTTGAACCTTGGCAGGAGACAGCGCCCAACACGCGGCGGAGCCCTAAA
|-----|
237675
237702

231158 231179
231186 CCGTGATCGAGAGCCTGTTCAACCGCCTCAAGCCGCCACACCGCTAT
...
231253 AGGACTAAACCATGCACCCCATCGAAACCCCGACAGACCTTCCACGACGGGACGGCG
231313 TGTCCGAATTGGCACCATCTTACCCGCGTGGTGCTCAACCAAGT
...
231544 AGAGAACTGGAAGCGGGGCCAACAAAGCCTACATCGCGCGGATCGCGACAGCCACACA
231604 TCGTCTCGCGGAGGACACACCCTCGGCCTGCACAGCGCAACACCGCATTTCTACAAACA
231664 ACCACGA
...
231727 ACACACTGTCCGATACGGCATCGGCAATTTCAAAGTCGAACATTCCTGGGGCGATTGA
231787 ACACCTCAAACAGACGGCATCTATTCCCTGCGGAGCGGTCGGCAGCTCCAACCTGC
231847 CCCTTGAACACACTGCTGCCATATCCAAGTCATCGCCGCGCAGAAACAGCGGTGTGCA
231907 GGCAGTTGGGTTATCCGGCCTACAGTCCGACGTGACGACGCCACCAACGAGCAGCG
231967 CAACGACAACTGGTCCGCATGGAACAACTCAATTCGGACGCAATCCCGTCGGCGC
...
232167 ATCGGCATCACCGCATGGTTTCGTCGCGACCAATCCCGACCGCTGGCTGGCGTTTGAC
232227 GACATCCGCACGCGCTAACCAAAACCGGTTATCCGAGCTGTACCGTCTGCTGACCGGA
232287 AATACGGCAGCATCCAAACGTCCTCCGACGGCGGAAGACGGTTTATCCGCAACGCGCC
232347 AACGGCTTGGCAGTCGGAAGCAGGAGAGGAAATCAACGCGCA
...
232462 ACGAAAGCAGAGAAGCGGCTCGTATCGACTTGGACGCGGAAAAATTTAAACGACAACG
232522 GCTTTTAAACCCCGCCGCGACAGCAAAATGGCGACAGCGCGGCGACGAAAAACCGCCCA
232582 AAGCCCTGGTTTTAAACTGTGCATCAAAAGCCGCGACACCTTGGGCGGAAGCGGTGTTCT

232642 GGATAAAGTCCACGCGGAAACCGTCAACGCGCGCGCTGGACGCGGACACGCTGGCGC
232702 AAAACCTCCAAGACAAAGCAGA
...
232858 AGACCGCAGCTACAGGTTTACACGAGCGCGGCCCATAGAAAAAGAAAGTCTCTTCC
232918 CCGTCGCTTTTGGCGACGCAAGTCAAAATGCTTCTGATCCGAACGCCATTTCGGGACGCG
232978 TTACCGCGATCGAAGCAACAACTGGCTGTTTATCCGCGCAAAAAACACGCGCGCG
233038 C
...
233053 GGTACGAAGCAGTTTGCAGTGGTGCATCGGCAAGCGCTCGGAAAGCGCGCCA
233113 GCTCCCGATAGGCCCGCAATACCTGAACCAACGAGAACCGCA
...
233248 AGATGGCTGTAGAGATGGGTTGTAGAGACGGCTTCAGCCGCGATCCAAGCATCCGA
233308 CCGAAACCGCGCGCGCAATCCCGCGAAACCTATGCCCCGCCA
...
233636 GATAGTCGGCGCGCCGATAGCAGGGCGCGCGGTTATTTCATTTGGGCGACATAT
233696 TGGCGAAAAACCTCTGTTGATGCGCTGTGCGAGCCCTTGGCGCGCG
...
233900 GCGCGGCAAGCGCAAAATCTTCGGCGCAAGTTCCCGCACTTCGCGCTTGTGTT
233960 AAGGATTGAGCTTGCATATTTTAACTCTCTTTTATTTCGCTTTCGGAACATGATGAC
234020 ACGGATGCCGCTCTTTTATCGCGTAAACAGACAATGTGCAGCGCTTGGTATCGCTAG
234080 ATAAGCAGCGCAACATAACGCGGTTTCGGGTTAATCAAAAGCAGACAT
...
234206 ATTTTTTTCGGAGTCAAAATTCATCTTCATTCGAGCTTCGAGGTATTTTTCGTTTACAT
234266 TATAAGCGCAAAAAACCAAAATGTAAATACAAAAAAGGAAACCCCGAAATGACCATCTA
234326 TTTCAAAAAACGGCTTTTACGACGACACCTTGGGACGATCCCGGAGGCGGTTCCGCT

...
228232 AAACCGCAGCGGCTATTGTGGCAACCTGGCCAGCGTGTGCATGTGATCGAGCAGGAC
|-----|
228282
228292 ACGGTATCGATCGGCTGTTTTTCTGATGGGCGGCGTTTATGCT
...
228349 ATGGCAGCAAAACCGAGCTGCGGCTCAAAGAGACGGTATTGGACACCCGACGCTTACC
|-----|
228399
228409 CCAAAAAGCGGAGCGCGCGCAAGCGCAAAAGGCAAAACGCAAAAGG
...
228560 ATCGGCGAAACCTGCGCGGCGCTTTTCGGGGAATAATCACGCTGGTGTGCTCCGAG
|-----|
228610
228620 CCGATACAGCGCTGCAGTTGAGCGGCTTGGCGGAGCAAAACCTGCT
...
228773 TGCAGCCACGCGGACGTACCGCATCAAAAACCTTAAACCCGGCGAGACGCGGATTTT
|-----|
228823
228833 AATCATGAGGGTGCAAAAATCGTGATTAAAGCAAGGCAAAATCATTTG
...
228900 GGGTTAACTGCAAAATAACGAGTTAATGCGGCCACGGATGCCAAATTTAAACGCTCCGT
|-----|
228950
228960 TGGTGGAGACCAGTCAGTGTGACGGCGCAAGGCCAATCAACGGCAACGGCGGCATGG
|-----|
229020 CCGTCGAGGGCGCGGACGAGCCACCTTTAGCGCGCATGTTAAACCAACGGCGGCAGCT
|-----|
229031
229080 TTAAACACCGACGGCGAGCTGGTGGCGGCAATATATCTGTTGCGCCAGCACCCGCATACCG
|-----|
229091
229140 ACAGCATCGGCGCAAAACCTTACGGCGG
|-----|
...
229257 TCTGAAACGGCAAAATGCCAACATGGACAAAGAGCTAAACCCGCGACCGCGACTATA
|-----|
229307
229317 CCGCGCGCACCGTCGATACGCTGCAAAATGCCGTGTATATCCGTTTGATGACACCGTTGG
|-----|
229377 GCAGCTGTGGCGGATAAACGCTCGGCTCGCTGCTGCAATTTGTTGACGCGCAAAAG
|-----|
229396
229437 ACCTGCAACGGGTCA
|-----|
...
229599 ATTACCGCCACGAAGTGCCTGATTTAAAGAGGTTTTTAAACGTGTTTGAACGCGGACA
|-----|
229649
229659 TTTGAGCAAAATCCGCGAGCGTATCTGCGCGATACCAAAAGCCTGTGGCCGATGCCGAT
|-----|

229719 ATCAGCCCCAGACGAGCAACATTATGTGCACGCCAGCCGTTTGGCCAGCTTGCGCCGAAGG
229731
229779 CAATATGC
...
229916 CGGTATTGGCAATCCATGCTTTTCAGACGGCTTCAAGTCCGTATCGGCCAGCGTTTAA
|-----|
229966
229976 CCGCACATACCGCCCGCGCTTATCGGCAGCGCGCACGGGAAATACCGGCAATCGC
|-----|
230031
230036 CGACGAGCGGCGGCGCGCCCAATGTGCGCGACGCGGAGGCGCAACTGAT
|-----|
...
230101 GTGTGTCCACGAATGCCGCCCTTACCGTACAAGGGCGGACCCAGCAGAGAAAGGATGCCT
|-----|
230151
230161 CACTGTGGCGCTGTGTTGAAATCATCCGCGACCCGCCGAGCGGCAACCGTTACG
|-----|
230221 ACTATAAAACCTGGCGTTGAGTGTGACGGGTAAACGAGCATATGTTTATCCGCTGC
|-----|
230234
230281 GCCGCGGCT
...
230432 GCCTGTTACCGTGCAAGTCAAGCTCGACGGCATCGACTTGGACGAGGCCAGCGCGCAT
|-----|
230482
230492 ACGGACGGCCCTAAAGAATAATTCGACACCCCTGATCCCGCGGCGGCTGACTGTGTC
|-----|
230532
230552 GCAATCGAGCGGCTATCAGCAATGTGGATGGTGT
|-----|
...
230738 TATGCCCGCAATGCCCGCGTGTGCGGGCGCAGGAGAGAAATAGACGGCGCAGCGTGGAT
|-----|
230788
230798 GCGTGGCGGAATCGGCTCAAAGCGTTGCCGATGCCGTGACCCAGCAGCGCGGCCAA
|-----|
230819
230858 ATGCTGGCGGATGGGAGCGCGTATTAGTGTGGACGGTACGGGCAAAACCGCCAGCGC
|-----|
230913
230918 CGTGTGTTGGCGCTCATGGCCAAAGCTAAACGAAACAGCGGCTTGAGTATT
|-----|
230918
...
231006 AAATCCAAATCGACGAACCGACGCGTTCGCGCGCGGTGTAACCCGCGCGCGACCGTC
|-----|
231056
231066 TTGCGCCGAGGAAATCATGTGGGTGTGGCACGTTAACGTGCGCGCGGCAACACCGCA
|-----|
231126 TTATCCGATTCCGCGCGGTATCTCGGGCGGCGGACAGCTGACCGATTACAGCGATG
|-----|

225458

225493 TTATTTTCGGGTCTGATACCGTGAAATTTATAAAACCTCAA

...

225557 TTTTATCGGGTAATACAAACCCCTGCCAAGATATAAAAATCAANTCCTAGACGCTTCTA
|-----|
225607225617 AAAAGCCCTGAAAACGATTAAATTGTGTATCGCGCGACAGGTTTAAAAAATGGCGG
|-----|
225630

225677 AGGTTTGA

...

225807 AGAGGGCGAATCCAACCTGCTGCCATATGCGGAATTCGCGCAGTAGACGGTCGTCCGAC
|-----|
225857

225867 TGATGTCCCTGCGTGTATCTGACCGAAGAAAACGGTTCATGATGC

...

225916 TTGTTGGCCAAACGCTGCGCAATCAGTTGGTTGTGATTTATGAACACTAGACGCTCTAC
|-----|
225966

225976 AAAGAGAAAACGGACANCTGCACCTGCGCGCGTTGGATGCGTT

...

226053 CGAAGTGGAGTGCAGCGACAAGCGGCTGCGCAATTGCGGCAAAAGAGTATCGCTACAT
|-----|
226103

226113 CTCTGCTGTGTTTTCCTATGACACAAAGGATATGTAAGCAAAAT

...

226168 GCCTGACAAATTTCCCGGTTGGACGGTATGGACAGGTGCTGGCGCAGCGTCGGCG
|-----|
226218

226228 CAAATTTAAACCGAAACGGAGCAAAACCCCTATGAAAGAGTTGT

...

226445 ATTTGACTAAATACGGCCCTATCTCAGTGGTTCAAGAGCTGCAAAAGCAAGTCGCGCGC
|-----|
226495

226505 TGACTGCCAAGCAGGAAGCAGACAAAGCAACGAATTGATTACCGC

...

226601 GCGTATTGAAACAGCGCGGCGCTTGCCATTTTGTACCGGCTTTATTGAAAACGCCCAGC
|-----|
226651226661 CGTCTGCTGACTTGGCAGGCTCGCAAAACGGGCGGTAAAGCACCCGACAAACGCGTCGCCG
|-----|
226663 226714

226721 CACTGACTCGGAAGAGCAGCGCAGCAAAAATGCTGGGCATGTCGG

...

226770 CGAAGAATTTGTAATAAATCAAGAAAGCGAAGGTAAGTAATGGACAAATCAGCGATTTCG
|-----|
226820

226830 ACCGCAATCAGCGCAGCATTCGCAAGAAATTTCAACGGGGCCT

...

226901 GCCAAACAGGTGCAGCCGTAGCCAGCGGTGCGCCTGTTATCCACGTATCGCTGTTG
|-----|
226951

226961 TCGGTGGCAACGGCATTAAATCGAGCGCATGGCGAAGAGATGACCG

...

227067 CCTTACGGCGGTGCAGACGGTGTCTCCGAGTCTGTGGGCTGACGGCCAAACGCCCTGT
|-----|
227117

227127 ATACCGAGGCTTACCAAAACGGCAGAAATCCCTGCGCGCGCGGCGAGG

...

227211 AGCCGCGCTGATTGTGCGCAAGCCCAATCGACGGTACGATACACCAAAATCGCCACG
|-----|
227261227271 AGTTTACGGCGATATAGCCCCGCGCAGCAGAGCTGGTGCGGCTCAATCCCATATCCACC
|-----|
227279

227331 ACCC

...

227387 CGGCTATGCGGTGCTGCGTGGCGGTAAAGAGCACCCTGAGGCGCTACGA
|-----|
227437227447 CATCGACAGCGACTTTTAAATCCCTGCCGACAGCTTCGATTTGTATCGCAGGTTGGG
|-----|
227454

227507 GCC

...

227546 CGAGTAGTGATAGACGGCAAAATCGTGATAGGGGCATCATCGGCAGCCGCGCGCACATTG
|-----|
227596

227606 CAAAGCAAGGCGCGCGCGAGTTGAGCTTGAGGGGCGTGATTG

...

227877 CCGGACGGCACGTTGTTGGTGGCGGTGTGGATTACAGCAGCCCGGGTGGGACATTG
|-----|
227927

227937 TGTGGAGCGCACCGCAGCAGCCCGCAATATCAGCGCATGGACA

...

227998 ACAACCGCTTTCTGAGGTTACTTTTTTGGCGCAATGCGACGCCCGCGCAGCGCAGCG
|-----|
228048 228054

-| |-----|

228058 CCAACACGATTTAAAGTGGGTGTACAAAGACCCGACGATGACGCTGCACCCGCCATAAA
|-----|
228092 228098

228059

228118 CGGTGGTGTGCCGATGCCGACAATTTGGCCGCAT

```
222395 222411
222421 CGCGAAATGCGACGCCCGGAAGAAAGCTGTCCGACCAAGCCTACGAAATGATGGACA
|-----|
222431
222481 GCCTGC
...
222627 GACAAAGACAACGGGCTGCTGCTGTACCCGGGAAATCCGAAGCGGAAGCGTGTGG
|-----|
222677
222687 CCGCTGGGCTGGGTCGCTCATACCCAAAATCCGAGCTCCAGCAGGCGCGCAACGGG
|-----|
222699
222747 CTTTCCGCACGCTTTCCCTGGCTGTATGTT
...
223002 CAGATGCCGACTGGTTCGAAAAATCGCGCGCGCTGATTTTGGGGCAACGCTAAC
|-----|
223052
223062 AGCGGTGCGGACGGAAAATCCAGCACCAACGCGCTGGGCAATATCC
...
223231 CGAAATTGAAATTTGACACGCGGAGCCGAAAGACATCGCGGTCTTTGCGGAGCTATCC
|-----|
223281
223291 CGAAACTGGTGGATGTCGGCGTACAAATCCCGAAAGCTGGGTGCG
...
223624 ATTTGGACAACGGAAACTCGGTACCTATATATGACGAGGCCCTGTTTATCAGCGATATT
|-----|
223674
223684 TGGGACAAGACCATGCCCCGCGCTGATTTGGGATTTGCCTTAAGTC
...
223808 AAATTGCCAAAAGTCTATACGATTTGCCGCGATGACCGACTTTGGATATGCTCAACGACATCA
|-----|
223858
223868 AAACCTCGATGGTTGAATCGGCAAAAAGTGGACAGTCGTTGACGATTTGCCGAAAAGGTA
|-----|
223902 223909
223928 TCTTGAATCTGCTCAGCAACAAGGGTGGCTGCATCC
...
223987 ATCATCGACCCACGCCCGCGAGGTATTGCGTTCGCCGCGAGGTGGAGACGATTTAC
|-----|
224037
224047 CGTACCAACATGCAAACTGCCTACACGCCGGTCAATATCAAGGATATATGGCAATATT
|-----|
224071
224107 GATGCACGACCTTATTGGATGTATGACGGGTAGCGACAGCCGCGACCCGCTCCGCGCAT
|-----|
224140
224167 TCGGCAATAGACGGGCTGGTGTACCGCTACGACGACCCGTTTGGGCAACGTTTACC
|-----|
224197 224214
```

```
224227 CCCAACGGCTACAACTGCCGTTGCTCGTCTCATCGCGCTGTCGGAGCGGATGTGGAAGC
|-----|
224282
224287 CAGGGCGGATTTGTCGGCAAGCACGTCGGACAATCTTGTGTAGACCCATAAAAATCTAC
|
224287
224347 AACAAAAAGCGGATACTTATCTGACCCCTTGCCTATATAAGCACGGGATGGAGTCTGTAC
|-----|
224356
224407 ACGNCCGATCGAGGATTTGATTACAGCGCCGAGCAATGAATACCCGCCGATTTAGAC
|-----|
224431
224467 AAGTACGACCGTGC GTTGGC
...
224523 TTTTAAACAGCCTTTAAACAGCTTGAAAAGAGTTTTTATGAAGTCAAGCAACGTTTGG
|-----|
224573
224583 TATTGATGGCAAGCCCGATAAAGAGCAGAAAAATCAAAAATCCGAAAT
...
224653 GCTGCGGGTGTATTGAGCAAGAAACGAAAGAAATTGGCAGGTATGACACGAGCGACGGTG
|-----|
224703
224713 TGGCTGCTGATGATACGTTGGTTAAACAGGTAGACAGCCGCTGAAG
...
224891 AATATATTAGGAGGTGGATGAGATTTATCTACAGTCGTACCGNAATCAGTAACGACAAG
|-----|
224941
224951 AGATTGCCAAATTTATGGCGGAAGAAGAAAGTATTGAAATAGACGTTGGGCAAGGCTGAA
|-----|
224991
225011 ATCACTTGCACACGCTCTCGGACGCCCTAACGGGCGAGGCTGCGGTATCGAGGTTATCACC
|-----|
225031
225071 GCTTTTCCACGTCGTGTACGGAATAATACCATGATTGATGTCAAAATAGACAATATCTTT
|-----|
225079
225131 GTCGTCCTAAACCAAAATCGAGCGCTTGGCAACGGGATCGAAAAACCGTACCTGCT
|-----|
225131
...
225253 GAAATGGTTGGGCTTAAATACCCGACGCGCAAGCCGCTTTCGGATTTCGGGTGCTGTGAA
|-----|
225303
225313 AGACAGTTTTTCCACACTGTTCAGACAACGATACGCCCTTGTGCGGTACGAATATCGTCTA
|-----|
225338
225373 TGCCGCCATCCACAACCTTCGGCGGTATGGCGGGCGCAACCGCAAAAGTTCGGATTCCGCA
|-----|
...
225433 ACGGGAATTTTTCACGCTGACGACGACGACAAACAGGCTTTTGATGGAGTATGCGAGGA
|-----|
225445 225455 225479
```

```
219806 TTTTGAAGTCTTGGCGCGTATCGAAGCCAAAGCAGGATGACATGCT
...
219898 GACTGCAAGAAATCTGCGGCGTTTACGGCGGTCTCGCGGTGTGATTGTAAGCAGCGGC
219958 TGGGAGCTGCTGAGAGCCAAGTTCCGGGGCTGATATGGCCCATCCGGAAGAAATCCGCGA
220018 AAAGCTGCGCGGATTTGATCGTCAGCGACGGGAGACGCTTTGAAATCGCGCGGATGATGTG
220078 CGAAATCCCGATTTCGACCGCCGCGAGCTGGAACGCGCGCAAGAAAGACGACGACGA
220138 TTGGGACAAAGTGGCGCCGCTATACCTTGGCAGCGCGCGCATCGAAGACTTGAAGCCG
220198 TTCGCTGCTGGCGGTTTTTTTGGTGAGTACCAGTCAACGATGACAAATGTTGCAAGACAC
220258 GTCGATTGAAGGCTGATCGCGTCCGAGCGG
220290 CCAAAGTGTGACAAGCCTGTCGAGCGGTTTACCAAGACCGTGGCGGCAAGCCCAAG
220350 TGATGCGGNAACGTCAAAACCTGGCGACGCAATCGAGGTGTTGGAATGTTCCGGCGAAG
220410 TGCTCAGGAGCGATACCCCAACACTTGCAGGCTTATGTCGAGCTGGTCGAGCCGTTGG
220470 GTGTG
...
220558 GCCGCCCACTCCGCCAAATTTAGGCGAGAGGTGGACGGCTTCGATGCGTCGCCCAAG
220618 GCTATTGCTGCAGCGCGTGCCAAAGTGTGTTGACCCGCTAGCGGTT
...
220979 TATGCGGGCAGGCCCGCGTATGCGAGCGCGGTACGATTTGTGACGGCAATGACGTTAAAG
221039 TCCAAGTGGCGGTAGCGGTAAAAGCTGCGCGTTTGGCTCAGGCGCTTACCGTCTG
221099 ACTTAAGTGTGAGCGATATTGAGAAATGACGAGCAAGTCCGCAACCCCGAAGCGCGG
221159 ACAAGCTCAA
...
221260 GTACTTAACCGCACTTTTGAATAACCCGTTTGGCAGCGTATTAAAGTTTAAGCGCATGAAA
```

221310 221318

```
221320 CGCTGGCCCCGACCGCATGGATTTTGGGACAGATGGGAGGAACCTTTTCCGAAACGACGGC
221380 GAGACTGTGGCCGAGCGCTTTTATCTCGCCCAATAAAGACGAGATGGAGCGCGGTGCACA
221440 ACAAGCTGGG
...
221474 TGAATAATCCGTGGCGCTGACGGTCAATGCGACATTTGACAGCAATATCAAAACGACCCGG
221534 TCAGCGCGGAAGATGCCCGCTTTGCCCAAGTCGATGAAGTTTGGAAAGGACCTGCCGTCCG
221594 ATTTGGTGTATTTTCGGTGGCTCGACCCGCTCGGTGGGTAAAGCGGGGCGAGCGGTGACC
221654 CGTCGGCAATCGTCAATCGCGCGCTATCAGCGGTCAATGGCAGGCTTTATGTCGTGGAGG
221714 CGCAATCAAAAAGAGTCTGCCGACCTGATATCGAAGAGTTATCTGCTGCACCGTC
221774 AATATCGTTCAAAACCTGTGTTTTCGAGACGCTGCAGTTTCAGGAGTTTTTAAAGAC
...
221923 TTGCAGCCTCACATGGCAACGGTTTGATTCTGTTGAATGAGAGCCAAACAGACGCTGATA
221983 CAGCAGTTCGCCCATTTTCCAAAGGCTGATCATGATGATGTCCTG
...
222050 CGGGGGGGTGGCCAAATTTGTGCCGATAGAAATGGCAAGCCCTACCGATAACGATTTTG
222110 ATGACGAGATAAAAAGTAAATGGAGCGGATAATGGCAAAAAAGAAC
...
222165 AAAATCCAAAAGCCGAAAGCTGCATTGCAGAGGACGTGGCTCAAAATACAGGACCGGT
222225 CGAGTTATCGCCGAGCATCCATCCAATTTTATTACGCCGCAAAAGATGCGCGCCTCTTC
222285 GA
...
222301 GTGACATCCGCCGCCCAACACAGAGCTTTTCGGGACATTTGAGAGCGCGACGACATCG
222361 CGCAAAATATGGGACCGCAAAAGCGCGTGTCTGACGCTCAACTGGCGGCTCGCCCGC
```

217161 TTCAGACGGTCTTTTTTTTGGTTTGCAGGGTGAACATCTACCGTTCCGGACGATGGGT |----|
217211
217221 TAAAGACGGTTTAATGGGTTTTTCAAAATGTTATAGTGTATAGTGTAAAGTGTCTGAAGCACCAGTGAATCG |----|
217304
217281 TCGCCTGGCCTTAGCTCAAGAGACGATTCCTAAGTGTCTAAGTGTCTGAAGCACCAGTGAATCG |----|
217282
217341 GTTCCGTACTATCTGTCTGCGGCTTCGTGCGCTTGTCTCTGATTTTGTGTAATCCA |----|
217372
217401 CTATATGCAATCAATATTTTTTGGAGA
...
217447 TAACCTTAACCGCAGCCACAGCAACACCGACCCGGTGGGTCAACGGAAGCGACCGTG |----|
217497
217507 AGCGGACTTGGCGCAGGATATGCGCAATATCTGTGCTGCTATTTT
...
217599 GGCATATACCGCTGCGCGAAGCGGTCAAGCTGATTCGGGCTCGGATGTGGCGATTGAG |----|
217649
217659 TTTCACACCAACGCTCCCGTCAGCAAAAGCGGACAGGATCGAAGCCTTGAGTACCGTT |----|
217688
217719 AAAAAACAACGCTGTGTCTCAGGTGTTGAGCAAAAGCGTTGCCAAGAAACCGGCTGAAA |----|
217726
217779 CTG
...
217831 TGGCTTATGCACAAGCCGGCGCATTTGTGTTGAGCCTTTTTCATCAGACGACACG |----|
217881
217891 ATTTGGCCTTGTTTAAGACGACTAAATGGGCATCTGCCGCGGATTCGGGACGCGATTG |----|
217907
217951 CGATGGAATTGG
...
218122 AAAAACTCAGCGGTGGCAGCCCGGAACCGCGGCGGCGCAACGAAGATCTCGTCTTA |----|
218172
218175
218182 ATCATGGACAGTTCCGTTTTACGAAAGATGATTGGCTGTGTTGCTCGGCCAAAGCCGT |----|
218268
218242 GAGGCAAGCAACAGCAGACTTGGCGGCTTTGAAATTTTCGGAGGGCGAATGGCCAAT |----|
218268
218302 TATCAGGAATGCTGAACGGC
...
218331 AACCAAAAGTGTGGAATAGCAATGCGTATTTTGGATATTTTAAAAACCCAGCAGCGC |----|
218381
218391 AATGTGTCGACTCGAAACTGTGGGCAAAACGTTTGCCTGCGCGCGGGGAGGTTAAGTTT |----|
218418
218451 GTGATGCTGCCCGACCCGTCGGC
...
218485 CGGTGATTTGGGCAATTGTGCGCGGCTATGCGGTGGCGCTTCGTTGGTCAGCGTCAAAC |----|
218535 218542
218545 --|
218547
...
219020 TTGATGTCGCGCCCGGTACGCCAAATCCGCCGAAAGACGGAAGACAGTAACGCTGTG |----|
219070
219080 GAACACGCCCGCGAGTTTGGCGGCTATGTGGCGGAGTTGGAAAAATC
...
219142 CGACTGGGTCAACAGTCAAGCGCAGTTGACGCTTCGGAGGTTGCGGATGACGACTTAT |----|
219192
219202 CCGTGTGTCGCAACGACGCTCGCCTGCCGCCATCGCGGATTTAGAGTGGGATTGAGC |----|
219223
219262 CGCGCACGCGAGCAAGA
...
219307 TTGGATAAGCCCGTATCGATTACGGGTACGCATGGATAAGGATTTTCAGCAGCAGCTTT |----|
219357
219367 TGTGTGGAGTTTCCGCGACGCCCGCTGCTGATGTGATTTGCGGAAATATTACT |----|
219384
219427 CAGTCTTTTCAGACGCCCAAAAGGTCGAGCGCGGAGTCGTCATCCCGAAGGCTACGCG |----|
219463
219487 TCCGTATCGTATTTGGCGATGTGCCGTTTAAAGGGTTTTAAATGGATTTTGAATTTGG |----|
219501
219547 GTTTAAACCCCTGTGGCCGATTGCAACGGCGGCTTTGGTTTGGGTCAACGGCATTTTC |----|
219577
219607 AGGCCGCTCTGAAGAGCGGCAACAGCGCATTTGACGACCTTAAGAGGAGCTGCACCGGT |----|
219638
219667 CAAGCTCTTTATCATACCAAGCAAGA
...
219746 AAGTTGGAAAAAGTAAACGAAAAATTTGGACAGGAAAGCGGACAAATCATGACGACCCGA |----|
219796

218331 AACCAAAAGTGTGGAATAGCAATGCGTATTTTGGATATTTTAAAAACCCAGCAGCGC |----|
218381
218391 AATGTGTCGACTCGAAACTGTGGGCAAAACGTTTGCCTGCGCGCGGGGAGGTTAAGTTT |----|
218418
218451 GTGATGCTGCCCGACCCGTCGGC
...
218485 CGGTGATTTGGGCAATTGTGCGCGGCTATGCGGTGGCGCTTCGTTGGTCAGCGTCAAAC |----|
218535 218542
218545 --|
218547
...
219020 TTGATGTCGCGCCCGGTACGCCAAATCCGCCGAAAGACGGAAGACAGTAACGCTGTG |----|
219070
219080 GAACACGCCCGCGAGTTTGGCGGCTATGTGGCGGAGTTGGAAAAATC
...
219142 CGACTGGGTCAACAGTCAAGCGCAGTTGACGCTTCGGAGGTTGCGGATGACGACTTAT |----|
219192
219202 CCGTGTGTCGCAACGACGCTCGCCTGCCGCCATCGCGGATTTAGAGTGGGATTGAGC |----|
219223
219262 CGCGCACGCGAGCAAGA
...
219307 TTGGATAAGCCCGTATCGATTACGGGTACGCATGGATAAGGATTTTCAGCAGCAGCTTT |----|
219357
219367 TGTGTGGAGTTTCCGCGACGCCCGCTGCTGATGTGATTTGCGGAAATATTACT |----|
219384
219427 CAGTCTTTTCAGACGCCCAAAAGGTCGAGCGCGGAGTCGTCATCCCGAAGGCTACGCG |----|
219463
219487 TCCGTATCGTATTTGGCGATGTGCCGTTTAAAGGGTTTTAAATGGATTTTGAATTTGG |----|
219501
219547 GTTTAAACCCCTGTGGCCGATTGCAACGGCGGCTTTGGTTTGGGTCAACGGCATTTTC |----|
219577
219607 AGGCCGCTCTGAAGAGCGGCAACAGCGCATTTGACGACCTTAAGAGGAGCTGCACCGGT |----|
219638
219667 CAAGCTCTTTATCATACCAAGCAAGA
...
219746 AAGTTGGAAAAAGTAAACGAAAAATTTGGACAGGAAAGCGGACAAATCATGACGACCCGA |----|
219796

213429 AATGCATCGATCGGGTATAGCAGGCGATCAGGCAGGGCTTTTGCCGTGTTCTCCATTTCG 213419
213489 ATAAAATATCGCGGGCTTTGGGCTCCTTTGTGCTTGGCTCCACCATGCACAGCTCTTTT 213508 213518
213549 GGCATGTCGAGGGATAAATAATACC 213508
...
213629 GTCCAGCCGATAAAAATCAAGGCTTGTGTAAATCCGTATTCTTCAATGCGAGCGCTGAAT 213679
213689 CCATTAGAAAATGGTGTTCACACCTAAGAACTTATGCAGGTCA
...
213800 ATTCAATGGTAAATATATCTTTTCATTTCAGGTTGAGATGACAAGGAAAAAGACGCTCTAA 213850 213858
213860 CGTTGCTTTTCCACTATCCCGGATANGCTGTCACTTATCGGGGATTTCCATATC 213863
...
214041 AATCTAAATTTTCAGAAAATTTAAATCAAGTCTTTTCGGGAATTAAGCATACG 214091
214101 TTTTCACTACTCGCTTGTAGTGGGTATTTTCTTGTAGCCCTTTTAAAT
...
214472 TCAGCGGCTTTTGGTCAATCTTTCGCTGATATTCAGCTTGTGCTTTTAGCGTCTTTT 214522
214532 CATCAGGCATAGGCAATCCTTGATTTTAAATCGTTGTAGTAAATGCGTTGCAGAAAG 214563
214592 GTTGGCAAGGGGCGCGGTGTTCTGC
...
214895 TTCCGTTTCGGCATTTGAACGAGACTTCTTAATCACCCGTTTCTGCTCCTAAGCTTTTCG 214945
214955 GTTTTCTGCCAGTTTTTCATCAACCATAGGCAAACTCCTTGATTTTTTAAACCGTATTTT
215015 GTAAGATTGGGACTGGGGACAGCGCGACCCGCCAGCTGGTCAAGTCTTTAGTAGGCAT 215038
215075 TAGCACTTGCCCAACATGAA
...
215293 TTTACAGCGCATTTTTTGTGTGGGCATCAATGCCCTGCCATTTGTGCGGCGTTGCGG 215343
215353 TTATCAAGATATGGTATTTTAGTGTGTATTTTATAGAAGAAATTTAC
... |-----|

215859 GCAAAAACAGGGTGAACAGCAACATGGTGAGATTTTCAAAAGAGGTACACAAAACGCCGCTT 215909
215919 GAATGCCGTGATTGCTATTCTTTTCTGATTGCGCGGTGCTGATTACTCGCTTATGGCG
215979 CGATAAGGACAGCGTCTGCTATTGAAAATGTCAAAATCGCGGGCGGTAGCTTTGGGATA 215989 215992
216039 TTTTTCAGG
...
216074 TGGAAAAGCAACGGCGGAAAAAGACAGTATTAGATTGTAGTAAAAAAGGTCTCTGAA 216124 216133
216134 ACGTTTTGGCGAAACTCGCAAAAGCTCGTTTTTCAGACGGCATTTTATGCTCTGTC 216138
...
216241 ATATGCGTTAAATGGTCAATATATATCTGTTTATGAGTTTGAAATCGCTCGGGCTTGATT 216291
216301 GCGAAAATTAAGATTGCTCAAAAGGAGCTGGGCTTGATGACGGTA
...
216374 CGGATAAGCGTCTGTGCGGATATGGATGCTTCTGAACTTGAGTCTGTTGCTCTGATA 216424
216434 TCGGTCGCACGGGATTTAAGCCTAAAGCAAAAGGTAACCCACACGG
...
216569 CACACGAATGGCGCGCGGATGTTGGTAAGGATAGGTGCAATATTTAGACGATACGC 216619
216629 AGCTACATAAACTGTTGCTGCTGCGTTGCAGATTGCGGAAAAACAGGAA
...
216719 CATTTATGCCGATACCGTGTGGACATTTGTGATGTCATCGGACTTGGCAGCGCGGAA 216769
216779 CAGCTGTCNAGCGGATTTGCGGGCGCGGCTTTAAATTTGGTAAGGCAAGGTGGACACC 216790
216839 GAGCGTTTGGCAATTTTGGTTCGAAGCATCGCGGAAAGTGAACACATGAGCTGTTGCAG 216840
216899 GTATATGGTGGCAGGAATTTGATGTCCACGCTGGCGGCAAGGCTTAATACAGTTGAGA 216940
216959 AACCATAGGTTTTTATCAGGAGTTTGCAAAATTCGCGCATATTGATAGAAGACGGGCTT
217019 ATGGCATGACGAAGCTATGCCCTAAATACGGCATCTCTTTCAGCAACGGGATATACGA 217021
... |-----|

210026 AGCGGAGTATGCGGTATTGCCGAAACGGCACGACGATTTTGTGCTATTGTAGCGCG 210016
|-----|
210061
210086 ATAGATAAAGGAAACATGATGGACAACAGAAAAATGCGGCTTTTCGACCGAGTTGT 210125
|-----|
210125
210146 TGAATAATTGAACTCAAGCAGCCCTTGGCGGATTCACGATTTCAAGTAAAGATTCA 210184
|-----|
210206 AGGTGTTCCCGCTGACGGAATCAGGCTCAATT
...
210437 ACTTATGCGAAAAAGGCGTAATTATGAACCATCGAGCCGAAAAATCGAGCGTTTTTG 210487
|-----|
210497 GAGATGCAGGACAGTAACGGAACCGAAGAGCTGGTGTGGGTTTGTATCGACTAAG 210556
|
210557 ACGACCCCGGAATCGTGAAGTGATGCCGCTGCCACGAGGCGCGGAACAG 210561
|-----|
...
210710 TTACGGCACTTGCTGATGCGCAAGTTGGCAGTAGCGGCGAAGGTATCGCGGATGGGC 210760
|-----|
210770 AGTTTGAATGATTTGTTTGAATCGGTATCGGACAGATTGCGTGATT
...
211025 GACGAACCTTTTGAGATTCGAGAGCAGCTTTTGCCTGATGCGGACGAGAAGACGCTCTTG 211075
|-----|
211085 GRACTGTTAAACATAGCAACGGCAATACGCCCGGATGAGCAAT
...
211137 GCGCGCGGTACGCACGGCGAACAGACGGTATCAAGATGCAGCGCGGTATCGTTAAGA 211187
|-----|
211197 AATACAGTCCCTGATTATCCGATAGGCGTCTGAACGGTAAGTC
...
211369 TGTGTCGCGGTTGGTTGACGGTGGGCGTGCGGTGTTTGTGGCGGGAATGCGGTGCAAT 211419
|-----|
211429 TCCTGCGGAAAGAACCCGGAACCGGTTGCGAAAGACCGGACGAAG
...
211490 GACAGCGGATTTGGAGGTCAGGTATGGAGCGCAATACAGTCAATGACGCTCGAGG 211540
|-----|
211550 AAAAAATGNAAGGATTGTTTATGAGTATGAAGCCAATTTATTGA
...
|-----| |-

211605 CGAAAAAAGAGGCTTTTGATCGGCGGTCAAGAAATCCGTGCGAAATATGGCATAAAG 211655 211663
---| |-----|
211665 CGATTGTAAGGATGTGAAATGAGTTTCGACGACGTAATACGATTGGCAGGCTTG 211697
|-----|
211725 GGGACAACACCGCAGCGGTGCGACGCG
...
211858 GATTGCGGTGTTGCTGCTGGGATTTGGCTTGAATGCCGGCAGAAATCGTCTGT 211908
|-----|
211918 TGCACATGATAGCGAAGGCCATGACGGGTATGACGACAAATTAACAGGTAACCCGTTGA 211950
|-----|
211978 TGTGCTCTATATTTTTTTTGTCTTATTG
...
212131 CGTGCCTCGACGGGATGTGAAGACTGCCAGGCGGAAACTTTGGCTCAGGCGATAACT 212181
|-----|
212191 TTATGGAGGAAGTGTGATGTTTAAGTTTTTGAGAAACCCGCGGAA
...
212349 TCGAAGCGCACAAAGACAAAACGGAAGTTACTGCATTTATCCCCAAATTCGGCACTTGC 212399
|-----|
212409 AGCGCGAAGTCAAAACGCTTGAAACCGAAGCGGAGACAAAAAAGCGGCATCGAAGAAG 212421
|-----|
212469 AATATGCC
...
212676 AAAATCCGCTTTGCCGAGTTTGCCGCACTAAAAAGAAATCGACAAGACGCCATCC 212726
|-----|
212736 TGAATCAAAAAGAGCGTTTTCAGACGGCCAAGTGCCGGGATTAAGATTGTGAGCGCG 212748
|-----|
212796 TGGAGGAT
...
212976 CCTTTTTCATACCTAGATTGAGGCTGCTTCTCTCTGCTGCTACAGCGGCTGTAG 213026
|-----|
213036 AAAGCGCGACAGCTTGTGCGCTTTCGAGGCGGTTTCAATCA
...
213237 TCTGTTGCACCACCATTTTGGCTTTGGGACATATTCOAAGAAACCCAAATCGTTGAGT 213287
|-----|
213297 TTTTGTAGACGCTAGGAACCGCGCGGTTTCATGTCACAGATT
...
213369 ATACTGCGGTAGCGCAGCAGCGGAGCAATTTCCGGCGCGCTTGGAATAGGCGTCTCC
|-----|

206349 -| TTTTCCTGTGGGACTGTAGTTTTTGCACAGGAACAATAAGATAAAAAACCGCTGA
206350 206401
206409 AATCTTCGAAGACGTGGATTTCGGCGTTTTTTTGTATCCGGNAAGTTACGCCAGCTTT
206432
206469 TTCACAAAACCGCGCGGA
...
206555 AAATGGATTTAATCAATACTCCGGGCATACCACCTCAACAGCGCCTATGATGGCGATGTGG
206605 206611
206615 TCTTGGGTATTACTCAAACTATTTCAAACGGTGCCTAAGTGGATTTTCAGACGTTACA
206666
206675 ACCAGTTTCCCGGTATACGTTGCACAGTTTGACAAAGAGGTCATT
...
207068 TGTTTTACATTTAATTGATATAAGCGTATTCTTTTTCGCCAGTCGGGATCGTTGTTTC
207118
207128 TTTGATGTGCTGATTGCTCAACCTACCCAGCCATTTTCTGAGATAAGTTTCTGCTTTC
207193
207188 TCGTAATCGCTCAATCGGATTAGTTTGTAACTTGAACCTTTACAGTGACCATTCAGACTA
207193
207248 GCCCAAATTGCACGGATTGATTTTGGCGACCGTTTGGCAATCTCTCTGCTGCGCAACT
207272
207308 TGTTTAACAAAGTTGTGCAGCTTTGAAGCTTGCTCAAGCGTTTATATGTTCTTCCTGGCGGT
207344
207368 TTTGAATCGACACGGGTTCTGATGACGTGATTTTGGGTATTAAATTTGATGAACAACCGAA
207428 CCAGATGAGCGGATAACACGCTTACTGACGTTGCCGTTAATCTGATTTCTCAACGTTTGAT
207436 207445 207454
207488 TTGCCAATAGCGTCGTCCTCAATTTTTTATTAAAGTCTGTAACACGATTTAAATTTGCC
207496 207520
207548 GAATGAATGCCTGTAAGACAAATTGAA
...
208179 AGTAATAGGGGTGAGAAAAAAGCGCAAGGATGGTTGGCAACAGAGACGCTCAAGG
208229
208239 TAAGGTGGTGTGTGAATATGCTTACCTATTGAAATCCGAGCA
...
209966 CGGCTTTGGAACACGCGCGAGTTTGGGACGAGCTGGGACATTTGGCGGAAACGTCATCG

208309 GAGCTGGCGGAGAAGATGCGGAAAAACCTCTCCCCAAAGTCAGACCAGGACGGCGATGTCG
208359
208369 GCTCAGGCACTGGCTGAAGCGGCCAAGCTGTTGAACGAGAAACAAC
...
208630 TATTTGAAAGCGGAAAAACGCGAGCGGCTGTTTGGCTCCGAAAGACGACGAAG
208680
208690 GCGGTTAAGCCGATTGAGAGCTACGGTTGGTTGCCGATGTTTATGC
...
208867 CTGCCGCTGATTATGCAGGAGCGCGCAGGAAAAACGGGGCGGCTTATAAATCGCTGCTG
208917
208927 CCTTATGTAAACGCTGATTGGGGGCTTTGAAGCCGAACGATGTTTGGATCGCGGACGGC
208963
208987 CACAGCTTTAAGCGGAAGGTGGCGCATCCGGTACACGGCAGGCCGTT
...
209155 GGTTCGCCGATTATCTATTACTCGGATAACGGCGCGGCCAAACCGGACGAGATGTAC
209205
209215 CATGAAATCACGGGTATTACGTCGCCACTGGGTATCCGCCATGAAACGGGTATTGCGGGC
209300
209275 AACCCCAAGGTGCGAGGCATCATCGAACGATGGTGGNAAGACAATCTGATTGAGATGGCG
209346
209335 CGACAGTAGACACGTTTTCGAGCGCGGGGATGGACAGCAGCAGCAGAACCTGATGTAC
209346
209395 CGCAAGA
...
209602 AGGCTGGAACAGGACGCGTATCGCGCGGATATGTTCTCGCAGACAGAGCTGGCGGAGTG
209652
209662 TTTATGCCCAAGAGGTGCGAAAGGTACAGCGGTTGCGGTGGATTGT
209655
...
209732 CTGAGCTGGCGAGTATCAGAAAGACGAGGTACGGGTACGCTACGATTTGAGCGATGCGT
209782
209792 CGGCGTCAATGTTTGTATGATGAGCGCAAGTTTATCACTAAGGC
...
209850 CAATACCGCGAGGCTTTCCCGACGCGTGTATCGACCAACTGGCGGAAAAACGCGGAA
209900
209910 AGGCAAAATCAAGCGGGGAAAAATGCAATCAAGCTCGCAATGCG
...
209966 CGGCTTTGGAACACGCGCGAGTTTGGGACGAGCTGGGACATTTGGCGGAAACGTCATCG

202523 202523
|-----|
202530 CCGTCGGCTAGGTGAAAGCGGAGCGTTCAACATCAACGCCGATTTGGTAGCAGCA 202555
|-----|
202590 AATTGGCGAAGAAATTGAACGCCGAAAAAATCTTGATGATGACGAATATCGCGGTGTGA 202607
|-----|
202650 TGGACAAAACGGGCAATCTGCTGACCAAAACTCACGCCGAAACG 202637
|-----|
202695 TTGATGAACGTATCCCGACGCGCTGTATGGCGGTATGCTGCCGAAATCGCTTCTG 202745
|-----|
202755 CGGTGAAAGCCCGCTCAACGCGGTGAAGCCACGCATATCATCGA 202745
|-----|
202963 TCCAAACCGTCATTCCTGAAAAAATATAGTGGAATTAACAAAAACAGTACGCGTTGCCT 203013
|-----|
203023 CGCCTTAGCTCAAAGAGACGATTCCTAAGGTGCTGAAGCACCACCAAGTGAATCGGTTCCG 203040
|-----|
203083 TACTATTTGTACTGTCGGGCTTGGTCGCCCTTGCTTGATTTTGTGTTAAATCCACTATAG 203108
|-----|
203143 AAACAAAACAGAACCTTAAG 203108
|-----|
203218 GGGAAAAAGGTTCTTTAGATTTTATAGTGGATTAACAAAAACAGTACGCGTTGCCT 203268
|-----|
203278 CGCCTTAGCTCAAAGAGACGATTCCTTAAGGTGCTGAAGCACCACCAAGTGAATCGGTTCCG 203295
|-----|
203338 TACTATTTGTACTGTCGGGCTTGGTCGCCCTTGCTGATTTTGTGTTAAATCCACTATAC 203363
|-----|
203398 GTCCAGATTCCCACTTTCGT 203363
|-----|
204244 TAAATCCTGTGGCTTTTGAGTTTTTGGATTTACGCCCTCAATGCCGCTGTAACGCCGAAT 204294
|-----|
204304 CGGGCTCAGACGGCAATTGCGTCATTTGAAATTAACAAACCGGCCAGCCTTTTCTTTGGC 204304
|-----|
204364 TTCTTTTCCAGTCGGCATCGGGTTGACGGCGACGGTTTCGTCGACGAGGCGCAGCAG 204394
|-----|
204424 CGGCAGGTGCTTTTGGAGTCGCTGTAATAATAGGTTTTTGGCGTAGCTTTTCGAGCGTTTC 204430
|-----|
204484 GCCGGTTCGGGAGCCATTGGTTCAGCGGGTGATTTTACCTTCTTT 204476
|-----|

204626 CAGACGGGGGTGATGATGAACTCGTTGGTGCATGAGATGACGAGGGTTTCGTCGCTGCC 204676
204686 ATTTGGTGGCTCTGCACCAGCATACGCTGCATAGGCGAGATGTGGG
...
204842 CCGTTTTGGTAGTCGGGTAGAAATTTTTCGTTTTCGGCTTCGGTTTCGGCAGCTCAACC 204892
|-----|
204902 AAGCCTTTTTTGATGAGGTATTTGCGGCCAGGCGTGSTCGGAATCGG
...
205408 GCCTCGAACCGCGCATTCGGTCAGCGGCATGAGAGGGGCATCGATATAGCGCGTTTTCG 205458
|-----|
205468 CCGGCTTCGAGCGCGCTGTCGACCCACCGCTGCAAAAGATGGCTTTTGCCCCGACCTTCT 205477
|-----|
205528 TCACC
...
205676 GGTTGTAACCGGGTTTATGTACGTGTTTTCGCGTGCTTTATCAATTTGAACGATGCGG 205726
|-----|
205736 TCTGAAAACGGCTTCAGAGCGCATCGTTCAACCCGACGCAACGTTGCGGCACATCAGGGG 205758
|-----|
205796 CATATTTGGAACGCGTTTGAGCGTGCTTCGGGTGC
...
205881 AGCGCGGCTTCGGCGGTTTCGTCCAAAGCGCGGTTCGTCAGAGGAGCATGTTTCG 205931
|-----|
205941 GGCTTGGAAACGAGCGCGGACGAGCGCGCTTGCAGTTCGCCCCAGGAGAGTTTC 205967
|-----|
206001 TGCTGCCAGTCGTCGTGTTTGTCCAAATTTGTCAATCAGATAACCCAGCGCAGGTGTTTC 206009
|-----|
206061 ATGG
...
206093 GATGTCGGGATAACAAACCGCGTCGCCGAGGCTGCCCTGTGTCGGGTACGGCGTTTCGGG 206143
|-----|
206153 CAGGAAGAGGATGTCCTTGATGCGCGGACGCGTACTTTTTCGGCTG
...
206229 CGCAGCAACGAAGTTTTCGCCCAACCGCTCGGGCGCGCTATCAGCAGGGAATCGCGCTTC 206279
|-----|
206289 TTGAGTTTGTGATGATGCCGCTCAACAGGATTTTCGCGGTTGTGGCGAAACAGACGACG 206342
|-----|
206345 206345

197101 CAGCAGCATCCCGAAATGCGGAGGACCGCGCGCATCTATAATGGCGCAGCGCGG 197160
|-----|
197124
197161 TCCTGCTTTGGCGCACTGCAAAAGCTGACCGCGCTTGCCGAATCGCAAGGCA 197165
-----|
197165
...
197278 CAGGCCAACGGCTTATCATTTTCGACACACCCCTGGGGACGGGAAATATCGCCGGC 197328 197335
|-----|
197328 197335
197338 GCGGCTATGACCAAGAGGAAAAATCCGGTCGCGCGCGGATTTTCGAGGCGGTCAATGCG 197368
--|
197340
197398 GACATCATGAAGCGGTTTTCGCGCTGCGAGCCCGCGGATTTTCGAGGCGGTCAATGCG 197434
|-----|
197434
197458 ATTGCGAAACCATTACCAAAATCAACGGTCTG
...
197525 AGCGGGTGAGCGGTTTGGCCCGCGCGGAAAAACGCATCCACACCTTCATCGCCACCA 197575
|-----|
197575
197585 GCGCCATCCATATGAGTACAAATGAAGATGAAGCGGAAGCAGTGATTGAGCGCGCGG
|-----|
197659 197677
197645 TCAAGCGGTGAATTCGCTCGTGAATACACCGACGATTTGGAAATTTCTCTGCAAGACG 197659
|-----|
197659
197705 CGTTGGGTTGCGAAATCGATTTCTCTGC
...
197797 TATTCAATCCGTACAAAACCGAAGAAATTTTCCGCGAACTGATGTCAAAACGCGCAAC 197847
|-----|
197847
197857 GCGCGCAAGTCGTTTGGTGGCGCACACTGCCACACGATTTGGGCTTGGGGTTGCCAAT 197866
|-----|
197866
197917 TCGTTGCGCGCATTAAGGCGGTGCGCG 197890
|-----|
197890
...
197965 AACGGCTTGGCGAACGTCGAGCAATGCTTCGGTTGAAGAAATCGTGATGGCGTTGAAA 198015
|-----|
198015
198025 GTGCGCCACGACTTGTTCGGCTTGGAAACCGGCATCGATACCACG
...
198230 CGGTGCGCTGGGCAACAACCGTTTGAGCTTGGGCAAAATTTGCGCGCGCAACGCGCTTCA 198280
|-----|
198280
198290 AAACCAAGCTGGCGGATTTGGGCATCGAGTTGGAAAGCGAAGGACACTGAACGCGCGGT 198340
|-----|
198340
198350 TTGCGCGCTTCAAGAACTCGCGGACAAAAACGCGAAATCTTCGA
...
|-----|

198532 GAAAAACGGCTTCCGCCACAGGCTCCGCCCGCTGGATCGGATTTTCAAAGCATTTGAA 198582
|-----|
198582
198592 AGCGTGGCGCAAGCGCGGCTTTTCAGATTTATTCCTCAACGCCGTTGTCAACGGT 198634
|-----|
198634
198652 ACGAAAGCAGGCGGAAACCGCTCTGCGCGTGGCAACCGGTTGTCAACGGT 198673
|-----|
198673
198712 CAGGGTCGGATACCGACGTTTGGTCGCCACTGCCAAAGCCTACCTCTCTGCTTAAAGC 198727 198736
|-----|
198727 198736
198772 AAACCTGAATTTGGTTCTGCCAAACCGCAAGCGCAGCGGTACGATTTGAGCGTGA 198837
|-----|
198837
198832 AAACAGCATGCCGTCTGAAGCATAAAAAGGCTTCAGACGGCATTCGCCGATATAGTG 198880
|-----|
198880
198892 GATTAAATTTAAACAGTAGCGGCTTGCTCGCTTGCCTTACTATTTGTACTGTCTGCG 198912
|-----|
198912
198952 GCTTCGTCGCTTGTCTCTGATTTAAATTTAACTCACTATAATAGGGCGCAAAACCCATTG 198957
|-----|
198957
199012 G
...
199088 AGCGAACAGGCAACAGCATGCGTTGCGAGAGGTTCAAGGTTTATGACGGCGCTGTG 199138
|-----|
199138
199148 AGCGGCGCGCAAAATTGACACCGCTCGACTGGCTGCCCGAAGTGTGGCGGACGATCG 199196
|-----|
199196
199208 CAATTTACGCGCGCGAAGCTTCCGAAATCGAACGGCTGTTTTGGCAATGGCGATGGAA 199224
|-----|
199224
199268 ACACCGCGCGGATGTCGGATAAAACTCCCGATTTGTGGCTGTATGAAACGAGAC
|-----|
199334
199328 GCGCGCAGGATTTTACACATGGTGCAATGCTTATCTTTACGGTTTGGATATTGCGCG 199334
|-----|
199334
199388 ACCGATTGTTGAAGCCGTCGATGATGAAGCGTTTGAAGAGTTTATATCCCATCATG 199417
|-----|
199417
199448 GCATTGGCGGTATTTACGACGAA
...
199578 TCATCAACAGCGCGCAACCGTCCCGAGGAGGCGAAAAACAGCAGGACGATCCCT 199628
|-----|
199628
199638 GTCCGTGCGGACGGCGAGAAATACAAGCGGTGTGGGTGAAGATTGAAGCGTTGTT 199688
|-----|
199688
199698 TCCATGAACCAACGTAATAATGCCGCTCTGAACCGGATTTCCCATG
...
199756 TTTTTCACAGGCGGTGCTGCTGTTTTTTCATGCCGAACCGGACAAAGCGGATACCC
|-----|

193798 TGTTCCTGCCGCGGTACGGCTATCCCGACGGCTTCACGGCGC 193788
...
193864 TGCACCGCGTGTGCAATATCTGAACAGTTTGGCGAACCGACTATGTTGACGATATTT |-----|
193914
193924 TGAGCGGCGGTATGTCGACACGATTTGCTGGGAATCAGCGGAGCGCGACGCGGAACCG |-----|
193968
193984 ATCCGATGTACGACGAGCGGTGTCGGTTGTTTGAANAAC
...
194205 AGACTGTTTCAGACGGCATTTTATAGTGGATTAAACAAATCAGGACAAGCGGACGAGAG |-----|
194255
194265 CCGCAGACAGTACAAATAGTACGGAACGATTCACCTTGGTGCTTCAGCACCTTAGAGAT |-----|
194323
194325 CGTCTCTTTGAGTAAGCGGAGGACGCGGTACTGTTTAAATTTAATCCACTATATC
194328
194385 AGACATTTGAATTCGGATTAT
...
194571 GCAGAAATCGATATGGTTCGTGAAATAGTCCGACGCTGGCGGTTTTCGGTAACGATGCGG |-----|
194621
194631 TCTGAATGGTCGTGGCGCTTCGATGTGCCCAACCGCTCTTCGACCGAAGCGGACG |-----|
194684
194691 GTTTTCACGCGGAGGATGATCTATAAACTCGGTATCGAAATGCTCTGCACCCGCCGCT |-----|
194732
194751 TCGCGCGGATATGCCGCGCGCTGCGGATCCAAAC
...
194877 TGCACACGACGACTTCCATCGAGTTGCACAGGACGGGCGGTGGTTTGGCGTTGTAC |-----|
194927
194937 ACGATACGGAGCGCTTGTCCCAATCCGCTCTTGTGATATAAATGTGACAAATGCCCC |-----|
194946
194997 GTCCC
...
195144 AGCCGCTGGGTCGATGCGGGTTTGTGCCAACCCCGTTTTCAGGGCGGCAACGATGGCG |-----|
195194
195204 CGTGGGATTGGAATGCATCTTTGCCGCTGGGGAGTACGACCGCGCTGCCCGCTTTTCAGT
195264 GCCAAAGCCCGCATCGGACGTAACTGCGGGGCTTTCGTAATAATGCCGATAACG |-----|
195287
- | |-----| |-----|
195320

195324 CCCATCGCACGGCGCTTTTGGAGATTTCCAAGCCGTTGGGCAAAAGTCGAGGTTTCCAGT 195325 195332 195344
195384 ATTTGCCCCACGGGT
...
195675 TTCTAAAATTCGGTCAGAAAGCAGGCGGTATTTCCGGCGCTGATGGAATCCAGTCTGTCGCCGA |-----|
195725
195735 TGGATGAACACAGCCCTTTCGCCCTTACGCAATTTGAGCAGGCTCTTCGG
...
195903 AGGCTTTTTCCTCTGTTCCGACAAAGCGTGTCCGCACTTCGTCCACATAAAGCGCGCC |-----|
195953
195963 CTGCTTTCGGAATAGAACGCCACGCAATGCTTCTGCGTCCGCAAACTTTTGGCACGGGG |-----|
195978
196023 ACGAANAACGA
...
196072 CGGGTTTGAGCGAGGAACAGATATAGACCGGTACGCCGACTCGGTCGCAANTCGTCGCGG |-----|
196122 196125
196132 CTTTGATTTTACTCAGCATACCGCCTGTGCGCTTTTGCAGAACCCGAGCC
...
196264 TGTAAAGACCGCTATGTGCGTACGACGACCAAGAGGTCTGCCTGTATCTATCTCGCGCCA |-----|
196314
196324 CTTGCGCACTCAATGTGCTGTGCGCGATTTTCAATTCCTCAACCCGAAACGATCTCGT |-----|
196340 196346
196384 TTTTCATTTGATGGGAGCGCGCGGTTCGACGACGACGAAAGTGGCGCGCGGCAT |-----|
196408
196444 TTTGTTAGCGCGCTTTGTCGGCAAGTCGCGCGGCTGAGCAGGATTTTTCGGGACACGA |-----|
196453
196504 TGCCG
...
196530 TATTTCCATCAGCACCCCTTGCCTGAGCGGCGGGAAGCCTGTTGTGCGGATTTTG |-----|
196580
196590 ACCGACGCTTTTGAACCCAGCGCACCGAACCCCTGCGCAACCGCGCGGGAAGACACG |-----|
196594
...
197041 GATATAGACCGCTTGGTTGCTTATTTTCGGCGCGGTGAACGCGCTTTCGCAAGCGTTGAAA |-----|
197091

```
190138 CAGTTGGCAACAGGTTGCCCGTGGCGGAGACACACGATGCGGGAACGGCAGGTTGAGC      |-----|
190171                                     190171
190198 AGCCAAACGGCGTTGCAACGAGTGCCTC
...
190277 CCGTCTGAAGGTACCGCCATATGCCCGGTGCGNACTTAAGATAGCGGTTTGTCTCGAAA      |-----|
190327                                     190327
190337 GTGTTAATCCAATGGGCTTCAGTGCACAAATATGGCAGTTGAAA
...
190386 CGCTGAGGAAGGCTTCCGCCACGCCACAGAGATAAGACGGCGAGCGCGGTTCACACA      |-----|
190436                                     190436
190446 ATATTTCCGACGGAAAGCGTTTGCAGCATCCAAATACCAGTCCAGCAGCAGCCGGTCA      |-----|
190462                                     190462
190506 ACAGAAATGCCGG
...
190553 ATGAAAAATATTGGCGGCAGCGGTTGACCAGCATACGCGACAGCGCGGTTCAGCAGCAGC      |-----|
190603                                     190603
190613 GCGGCGAGAAATCAGCACCAGCGTTTCGGCGGATATCGGCCGGCAGAAATCGGGCAACAGC      |-----|
190633                                     190633
190673 AGCAGTAGGCGACGA
...
190748 ACGCGCAGCAGGTGGTAGTTTATCTGTGCAAGCTGTCCGCCCGCGCAGAGCGGTTTCAGA      |-----|
190798                                     190798
190808 CACCATGCTCGGAAAAAATTACCGGTACACAGTGGAAAGGCGGAAAC
...
191057 CGACCGGTTGTTTTTTTGGGATGGAACACAGTTAAAAATAAACCCGTTTAAATCGTTTGTG      |-----|
191107                                     191107
191117 TTGCGGAAACGGCATATGTTTTGGATAGTTTTTGATCGTTATTTTGTGCTTGCCTTGC      |-----|
191151                                     191151
191177 CGGCTTGTTTTTTTGTCCGCGCACAAATCCGA
...
191314 TGATGCTTTTCCATGCGGTCAAAACGGCAGTGTATTGGCTGTTTGTGCGTCTCGTCCGTT      |-----|
191364                                     191364
191374 TCTCCGAAACTATCTGGCGCACGAATCCGAACCGGACAGGCCCGT
...
191672 GGAAGAAGCAACCGCTGCTTTAAACAGCGCGGCTTTAAGGGAACGAAAAACGCTATAT      |-----|
191722                                     191722
191732 CGATGCAATTTAGAAAAACGAAACAGCGGTCCCAAAGTCCCGTGTCCGATACCCCGAT
```

```
191792 GGAAGGGTCGAGATTATCGGTTTGGACGACCTGTGCTTCAACGACGATATTCGGGTAT      |-----|
191817                                     191817
191852 GTTCGATCGGACAAAGAGCGTTTTCGAGTCTCGGAGTTACGGATTGAGCCGTATTT      |-----|
191870                                     191870
191912 TGAGAAGCAGCATC
...
192525 AGTGAGGGTGTGGCAGACGGGTCAAGCGGGCAATATTGTTCGGAACCGAAGCGTTCGGG      |-----|
192575                                     192575
192585 CATGACAGTCAGCGGTTTGTCCGTTTGAAAAATGTGCGCTCTGAAACGCCGTCGCCGCGG      |-----|
192628                                     192628
192645 GCATCGGATACGGAAGCGGATGAAGGCGGTTTCCAATCTGAAGAAACCGGTTCGGTATCC      |-----|
192671                                     192671
192705 GAACACCTGCCGACAAACCGACCTGCTTCTGCGCGCGTGTTCATATCCCGGCGACGCAA      |-----|
192755                                     192755
192765 ACCGAAGAAGAACTGTTTGGAAAAACAGCATCACCATCGAAGAAAAAT
...
192971 CGAAACCATCTCGGCACAAACCTGTATGGGTTTGGAACTTCCGAACCGCAACGCCAAAT      |-----|
193021                                     193021
193031 GATACGCTGAGCGAAATCTTCAATTCCGCCGAGTTTCCCGAATCCAAATCCAAGCTGAC      |--
193088                                     193088
193091 GCTCGCGCTCGGTCAGGACATCACCGGACAGCCCGTCGTAAACCGACTTTGGGCA      --|
193093                                     193093
...
193270 AAATGCTGGAATTGAGCATTTACGAAGGCATCCCGCACCTGCTCGCCCTGTCTGTTACCG      |-----|
193320                                     193320
193330 ATATGAAGCTGGCGCAACCGCTGAACACTGGTGTGTTAAGAAATGGAATAACGCTACC      |-----|
193381                                     193381
193390 GCCTGATGAGCTTTATGGCGTCGCAATCTTGGGGTTTCAATCAAAAAATCCCGGAG      |-----|
193440                                     193440
193450 CCGCAGCAAGGGGGAGAAAAATCGGCAACCCGTTTCAGCCTCAGCGC
...
193604 CCTCGCCAAAAAGCCCGCGGAGGCATCCATCTTATCTTGTCCACACACACGCCCGAC      |-----|
193654                                     193654
193664 TGTGATGTCATCACGGGTCGTGATTAAAGGGAACATCCCGACGCT
...
193738 ACAGCCGACGATTTCTTGACCAAATGGGTGGGAAAAACCTGCTCGGGCAGGCGGATATGC      |-----|
```

187041 TTACAGCCAGTAGTCCTGTTTTTCTCGTAAGCGGGTCGATGAG 187031
...
187193 TTTTGCAGATGACAAAGTCGGTGGGTGCGAGCTCAACACAGCGGTAATTTGTGCGCGACG 187243
|-----|
187253 CGGCTCAGCGATTTGTCAGGCCACCGAGAACCGCAGTAAGATTCTGGCGGAGGCGAGGATTT 187259
|-----|
187313 TT 187313
...
187325 GCGCAAAAGTCAGAGAGTTTCGGCAGCGAGGTTTTTTCGCCCTGTGCGCCTGTGCGAGCAGGGCGATGCGCT 187375
|-----|
187385 TGCCGGTATTGCGCGACTTTCTGCGCCCTCGCTGCTGCTTCGAGATTGT
...
187612 GGGAAACATGTTTTTCAGACGGCATGACGTAATGGAACAAATGCCGAATCAGCGTTTTTC 187662
|-----|
187672 GCCGATTTGCTTTCTTTGCTTTTAATCAGGTTGAGGATGTTGCTCTTATGGCGGAGCAA
|-----|
187732 CACCAATATGGCGATTTCGAGGGTTTGCAGAAATCCAGAAAGTATGCGGCATAAAAACAG 187741
|-----|
187792 TCGGGCAAGGGGGCGCGGATTTGCGGGTTAATGCGGCAAGGAGGACACCTTTGAAGCC 187807
|-----|
187852 GAATGCCATCA
...
188033 GGTTCCTTGAAGCAGCGGTGCAGCAAAACGGCAACCAACCTTTGGCGCATCGCCCAAG 188083
|-----|
188093 AGCGTCAGCGCGCGCGCTTTTTTTTGGCGCTGCGTAAACATTGGTTGCCCGCGGATTG 188093
|-----|
188153 CCGATCCGTAGTGTGCGGGTCGTCCCATGCGCTTAATTTTGACACGATGACGGCGAAA 188173
|-----|
188213 GAAAGTGAGCCGATCA
...
188370 ACGGCTGGAGCGCGAAGCGTTGCAGACCTGATTGTCGATTGGACATCGCGCTTCCCG 188420
|-----|
188430 AGAAACGGGTTTGGGACGACGATATTGCCAATACGGTGCTATTATGCCGAGGTATGCGAAA 188444
|-----|
188447 188447
...
188490 CGCTGCGCGACATCTGAAAGAACAGGATTTCTGCTTTTGGAAAGCGTTGGCGGAATATA 188493
|-----|
188499 188499

188550 TTGCGATTGTGTTTTTGGGATATTTTCGGCGGTTGGGTGCGCTGAAATCTGTCACGC 188600
|-----|
188610 CGGGTATTTTGAAGGCGTGCAGAGTTGGCGTGAATCGAGCG
...
188766 GCATCAGCAGGGATAAAGTCAAGTTGCCAACGGCAATGAAGGCGAGCTATCGTCATCC 188816
|-----|
188826 GGCATCCGGCGCGCGGTGCGCTTTGCGGTAACGGATGAGGAAAGGCTGTTTGGGTAC 188873
|-----|
188886 GGCAGTGGCGGTATGCGGCAATCAGCGGACATTTGGAACCTTCCTCGGGCAAGCTGGATG 188911
|-----|
188946 TGGCGGCGAGGATATGCGAGCATGCGCTTTACGAGAACTGGCTGAAGAACGCTTTATA 188995
|-----|
189006 CCACCGACAGCGTAGCGCTGCTTTACAGTTTTTATACGGCGGTG
...
189132 TTACGGAAACCGTATTGATGTCGAAGAGAAAGTCCGTCAGGCATTGGCAACAGATGAAA 189182
|-----|
189192 TTAAGACGGCAAGACATTAATCGGTTTGCAATACTAGGCTGATGAA
...
189524 CAAAAGGTTATGATGTTACTTTTTTTGCAATTTTAGACACAGTAATAGAGCTTTTTT 189574
|-----|
189584 TTAATAAAAAGCCTTGGTAGGTTTATCCTGCCAAGGCTTTTCTTT
...
189644 AGCCCAATTTGAATTTGAGAAGCGTATAGTGGATTAAACAAAAACCAGTATGCGCTGCGCT 189694
|-----|
189704 CGCCTTAGCTCCAGAGACGATTTCTCTAAGGTGCTGAAGCACCAGTGAATCGGTTCCG 189721
|-----|
189764 TACTATCTGTACTGTCTGCGGCTTCGTCGCTGCTGCTGATTTTTGTTAATCCACTATAA 189789
|-----|
189824 TTGGCACAGCAGTATTTAAA
...
189942 CAGCAGCAACCTGCCGTGTGCAGGAAATATTGGCAAGTGTGAAGCAGGACGATTGCA 189992
|-----|
190002 TTGGAGCAGGTTACGGTTTCCAGTGAAATCCGGAAGCGTGGTC
...
190078 AGCTTCCATTGCGGATGTTGACGTTTCGGCAAGATTTCCGATAGAAAAGCGCGCTC 190128
|-----|

183495 CACTTGATAATCAGATAAAATCTTGTAAGGATAATGACCGGGGTATTCCTTACCATTTTG
|-----|
183555 GACAACTGAACAGATATTTTGGGCATATTTACAGATATGCAGAGCTTTTGCACCTTGAGC
183564
183615 TGCTAACGCCTGAACACTGTGCAGCAAATTCCTAAATTTTTCAGCAAGCTGCCTGTTGTAG
183619
...
183713 GATAGCGTGCCAAATGCTGCTGATTTTGTCTCAACCGAGATGGCAAGCCTTTGGGGACACCT
|-----|
183763
183773 CAAGCGCTTGTTTGGCAAGCGGATCGGTGTATCAAAATAATTTTCCCAAGGCAATACAC
|-----|
183833 CGTATCGCTGATGATATTGTCTCCATCAGTTTAGGGATAGCAGCGGTACCCACCGAGCGAC
183836
183893 CACCGACCAACCGCTTCCATAAAATTCATAGGTTGACCATCTTTATCCAAAAATAATTCGG
|
183952
183953 GCCTCGACGCATCGGTGGGTCTCAGGCCCATCAAAATGGTGGTCAATGTTTGGCGGTAT
-----|
183957
184013 TATCCCAATACAAACAAATGCACCACCGCCCAAGCCTGCAGACTGTGGGTCTACCAAGC
|-----|
184051
184073 TTAGTGTGCTCTGCACCGCCACCATCGCATCTGCAGCGGTACCGCCTTGCTTTAAGATAT
|-----|
184078
184133 CATAGCCAGCTTGTTGCTAATGGATTGGCTGACGCTACCATCAAAATCACTTGCATCA
|-----|
184165
184193 CCTGCTTTTGTTCGGTCAGTCCCGTTGCATGTTTCAGGCGGTGAGCGCTCTGCACCTGTGA
|-----|
184236
184253 TGACAGCAGAATGAGTATTAAACCTTACCTTGATTGGCAT
...
184600 CCACCGATTAACTCACCNATTAACCCAAAGCTGTGGAAGGTTGGCCGATTTGGCGATGAGT
|-----|
184650
184660 GSTAGAGTACTGCGAATTTCTGGGTTTCTAGGATATTAAACAAAG
...
184764 GGAACCTGGGGCGTGCTTTCATATATAGTGGATTAAACAAAACCAAGTACGGCTTGCCT
|-----|
184814
184824 CGCCTTAGCTCAAGAGACGATTCCTAAGGTGCTGAAGCACCACCAAGTGAATCGGTTCCG
|-----|
184841
184884 TACTATTTGCTGCTGCGGCTTCGTCCCTTGCTCTGATTTTGTGTTAATCCACTATAC
|-----|
184909
184944 AGTAGGAAAGGCTGAAAAATTTATGCGTAAAGCGTGATATTGCAACGTTTTTATCAACCG
|-----|
184987

185004 GACGGCGGTGTTAAAGAAAAATTTTGGCGTATCCGATAA
...
185074 ATTAATGTAAGATTCAATTTGACTTTTAAACCGTAAACCAAGAGAGAAAGGATATGT
|-----|
185124
185134 TCCCAGAATACCGTGATTGATTTCCAAATTTGAACAGGAAAAATTCGCTTCGCCCGTC
|-----|
185194 TGTTCGACGAACACAAACAGAGCTGGAGGATAAAATTAACCGTCTGGTCAACAATCCGGTTA
185217
185254 CCAGCGGTGCGGAAACCAT
...
185410 GGAAGGTGGTTTCTATCTTTTAAGGCTATAAACGAAATATCGGCGGACGCTCGAT
|-----|
185460
185470 GCCTGCTGCCGCTTTCTTTATCGGTCAGGCTGCCTTTATTTTGAATGCAGCGGTTGA
|-----|
185530 GGTAAATCCACTTCTCGCTGCTGCCATATGACGGGAGCGCTGGTCAGGCTTTCGG
185564
185590 GACGGATGCCGAGTATGTCCTGACGGCGGTTGCTTGCCGATGCGCGGCTTGTTCGAGTA
|-----|
185615
185650 TCAGGCCATAGGATTGGCTT
...
186094 TTTCAAGGATGGAGAAAAATCGTACCGACGGAAATGTTGACATCAATATTGGAGATCCGT
|-----|
186144
186154 CTAAGGATCGAATAGGACGAGATAGCTCCGTTTTCACGGCACTTACGAAAGATCTT
|-----|
186178
186214 CTTCTCGCTTGCCAAACCGGCAACGGCAGGCTTGGCTTGAGTGTGTCATCATATGT
|-----|
186274 TGTTCGCGATTAACATCCAGTTTTTTTTTGGTCTTTCGCCCTGAATATTGCCCGTCCCGCCA
186278
186334 TACCCAATACGGCGCCAGTGCGCCGAGCGGACTTTGGCGTTGATTTCGGTGCAGCGG
|-----|
186371
186394 AAACAACGGACAGTAAACGCCGCCGAGTGCTTCGGGACGCTGGTTTGTTCAGGTTGT
|-----|
186410
186454 CGGGAGGAATC
...
186615 GGCAGCGGATAGGGGATATTTATTTCTTAAAGTCCCGACTGCCGATATATCGTCGGT
|-----|
186665
186675 TTTGTTTATTGACGGGAGATGTTGTCGAAGGGTTTCAGACGGCA
...
186981 GATACCAGATGAGATAACAGCCGACTCAAAACGTTTTTAAAGGGATTTTCAGCGTTTCG

181015 CCGATGCGGATCCGAGGTAACGAAGGTTTTTCCGGATCATTCGGGATGCGGTTTTTG |-----|
181065
181075 TTGACAGTGATGCGCTTTGCCAAAGCAGCTTCGGCGGCTTGC
...
181129 TTCAATCGGTTGAGGTCAACGAGGAAACAGTGGCTTTCCGGTCGGCCCGGAACGATGCGC |-----|
181179
181189 AAACCGCGTTTAAACCAACTCTTCGCCCATGCGCGCTGCAATTTTCACTTGTTTTGGC
181249 TATTGTTTGAACCTCGGTTGCAATGCTTCTTTAAACGCCACGGCTTTTGGCGCGGATAACG |-----|
181299
181309 TGCATCAGCGGGCCACCTTCAGAGCTTTGGGAAGATGGAAGAGTTTCAACGCTTTTTCGTGG |-----|
181354
181369 GTATTGTGCGGCGACAAAATTACGCCGCCGAGGACCGCG
...
181483 AGACCGCATAGTCGCCATATGCACAAAGAGTATGCGCGACTTTATCGCGCATTTTCG |-----|
181533
181543 CGGAATTTTGGCCAGTCGATTTGCAACGCGTAGGCACGACGCCGCCACAATCATTTTG
181603 GGTTTGTGTTTCGAGCGCGGCTTCGACTTCGGCATAATCGAGCACTTCGTTTTCATCC |-----|
181622
181663 AAACCATAACTAAGCAGCGTTGTAGAGTTTGCTGAGATATTAACGCTCGCGCGCTGGGTC |-----|
181704
181723 AGGTGGCGCGTGGCGCCAGAGACATACCCAAATGTTGCGCTGTTTAAACGGA |-----|
181761
181783 GCCTACAGCGTTGGTTGGCTTCGAGCGGAGTGGCGTTGGAGCTTGGCGTATTCGCGG |-----|
181824
181843 CCGAACAGTTTTTTGAGCGGTCATGCGCAATTTGTTGACAATATCGAGCTATTCGAG |-----|
181867
181903 CCGCGGTAGTAGCGTTTGGCGGGTAGCCCTTCGGCGTATTTGTTGGTCACTGGGAACCT |-----|
181913
181963 TCGCGT
...
181977 GGCGCAGCTGACGTAGTTTTTCAGAGCAATCAGCTCGAGCTGGTCTTGTGCGGTTGGTC |-----|
182027
182037 TTCTTGGCAATGGCTGCTGCCAAATCGGGTCATATTGTGCGAGGTTAACGCTTTTGA |-----|
182085
182097 AAACATGTTCTCGGCTCCTTTGTGTAATCAGGGTATCATGAGTG |-----|

182192 GTAAATTTTCTTTGGCATTATCAGGTAATTTATTTAACATGCTGGTTTTTAGCGTCTCAT |-----|
182242
182252 TACCTTTATGTTAGTACAAGGCTACCCAAAGCAATGCATCAATG
...
182400 CTGCCCCACCATGCTCTTAAATCAGATGAATGGCTTGTAACTGCTGTTTGGGCTCTCAC |-----|
182450
182460 AAATTGCGAAGATCTGCTCAGAAGCTTGCTTGATGGCTTGTATTTT
...
182715 TTGTTGCCACTGGAGGCGGTTGTACCGGTTTCTAGTTTCATAGCTACCAACGATTTAA |-----|
182765
182775 TAAATTGGTGCACCTGATGGCATCTTGATATCCATATTCAGTCACATGTGCCACAAT |-----|
182833
182835 CGCTTAGCGACATAGCCCAATGATACGCTACCACCTGGGAGCGGATGGCATATAAGG |-----|
182838 182846
182895 CT
...
182914 ATGGTTGGTGCATTTGAGAGCGTGGTCTCTTTCGGGCTCGACACGATTTGGCAGCTGT |-----|
182964
182974 TTGCCCTGCTTTTATTGGCTCAAACTAAAGTCTGTACGCTCATTTAT
...
183023 GCAGGTAGCCATTTGCCATCAAGTTGAGCCAAACGCAATTTTCAATGGAAGCTCGTCATTG |-----|
183073
183083 ATAGCACATTTCCCGCTTATCCACAATTGATATATGACTGGTAGAAGGTAACCTCATCG |-----|
183139
183143 CTTCGTAGGACACCCACTCATGAATAAATCGGCTGCAGATACGCTAGGCAATGCCTTAT |-----|
183144 183171
183203 CCGACTGCTCAAGCAGCTGGCTGC
...
183235 AGGTAGTCTTTAGAAATCAACTGGCGAATGGTACTGGTACAAATCAGGCTCGCCCAAA |-----|
183285
183295 TATACATCAGATCCGCAACGCAAGCCTAGAAGCGTCGCCCAAGAGACGATAACCTTCG |-----|
183327 183330
183355 GCATCATACCCACCTGATTGGGTGAAATT
...
183435 TGCACCCATACCGCATACTTTCATAAATAGTACACACAAAGCGGCGGTTCCAC |-----|
183485

173135 CGAGAGCGGATGGCGCAGCTTTGGCGCAACTGGGTGCGGATGCTCTCGAGGCGTTTCGGG 173185 |-----|
173195 GCGCGCTTCGTAGCCGTCGGGCTGCGACGGATCGCGGTAAACGACGAGATGGATTTTGCC 173234 |-----|
173255 GCTCAAAATTGTGATCAGCGCAATTCTTGCGACA
...
173540 GCCTTTGAGATAGTGGCTGCAAGGCGCCGATAAAGAATAGCGCGCAACGTTTCGCC 173590 |-----|
173600 ACCGACAACGGATTGAGCAGGTAGGTATAGGGCGGTTGGCGAGT
...
173901 ATTTAAACAGATAACCTGCGGTTACGGGAATCAGGCTAAGATGTTTGAACGTCCATG 173951 |-----|
173961 ATGCATACCAAGGTAAAGAACCGGCTGCGACCGATTGATTTCAGATA
...
174169 CGTACCGCAATCCAGGACCGCAAGTATCAGGACAATGATTTCTTCAGGTCTGTTTGAA 174219 |-----|
174229 CATTTTTTCTTCTGTTTGATTTTCAGACGGCATTCGCGTCTGTTTGGTTTCCAGCAGC
174289 TCCAGCGTTCCAGCTTTTCCAAAGCAGCATTTTCGATTTCTCGGCGCGCTTTGCAAT 174293 |-----|
...
174379 TCGGGATCGGAAAGCTCGCGGTGATTTCTGCCCTGTTCCGGCTTCCAAAGGCGCGATTTCG 174429 |-----|
174439 TCGGCGAGGGCTCGAGTTCCGCGCTGTTCTTTGTAGGACAGTTTGACCGTCCGCTTGCT 174447 |-----|
174499 TTGG
...
174703 TGATAATCGGCAGCAAACTTTCAGCAGCTCTTGGGTATCGATATCCAAGTCGTGTTGGTC 174753 |-----|
174763 GGTTCTGCCAAAATCAAAATATTAGCAGGACGGGTAAAGATTTTGCAGCAAAAGCGG
174823 TTGCGTTCGCGCGGAGAGCGGATGAAACGGGACTTTGTGCGCGGCTGGATGGAATAAG 174841 |-----|
174883 AAATCTTCCAGATAGCTCAATTAGGTGTTTTTCTTGCGCGGACTTCGACGTAATCGTTG 174936 |-----|
174943 CCTTGACCGAGCTGTAAACACGGGTCTGTTTTCTTCAACGTGCTGCGGAATTGCTCG 174969 |-----|
175003 AAATAGGCAACTTCTGCTTAC
...

175075 TTGAGGAAGGTGGTTTTTGCCGATACCGTTTGGGCCCGAATTAAGCCGATTTTGTCCGCACGC 175125 |-----|
175135 TGCAAGATGGCGGAGAAATTTGTCATGATGACTTTGTGCCCATAGTAAACGAGGCATGT 175170 |-----|
175195 TCCAGTTCCGGCAATGATTTTGCCGCTTTTCTGCGCGCTGTCGAGCTTGAATTGACCTGC
175255 CCTTGTACGTTGCGGCTTCGCGCACGCTGCGCGCGCAGTTCTTCCAACGCGCGCAGCGG 175268 |-----|
175315 COTTCGTTG
...
175453 GAGCCGGATAGGAACGCAAAATACCGCGATCGAGTTCGACAATCCGCGTTCGCGATATTG 175503 |-----|
175513 TCCAAAACGCGGCTGCTGGGTAAATCACACCAGGCTGCCCTTCAACGCTTTTGAGCAGG 175558 |-----|
175573 TTTTCCAACCAAAATAATCGCGTCGATGTCCAAATGGTTGGT
...
175615 GGTTCATCGACAGCAATACGTCGGCTTCTGCACCCAAAGCTGCGCAAGCGCGACGCGC 175665 |-----|
175675 TTTTCTGACCGCGGAGAGGTTGCGGATTTTTCGTTTTCCGGCA
...
175906 GTATCAATACGGTTGCTTCTCTTATCAAAAAAGGATTCTCGGGTACATAACGATTTTG 175956 |-----|
175966 AGGTTGTTTTGGACAATAATCTGCCATCATCAAGCTTTTGCACGC
...
176062 AAGCCGATTTTTCGCCCTGCTCGAGTTGGAGGAAGTTTTTGTGAGCAGGCGGACGTGG 176112 |-----|
176122 CCGACGGCAAGGAAGCGTTTCTTACAGATAATATATTATGATACAAAATCTCAACAGT 176136 |-----|
176182 TACCGTTTG
...
176350 CCGATGATGTTGTAGCGGATAAATTTGGCATGTGCAATTTTCCCATACCGCGCAAG 176400 |-----|
176410 GGGCGGAAGTGGCGAGTGGGCATAAAACGGGCAATGATGATGTTTTTCCGCCCTGT 176424 |-----|
176470 TTTTCGTAA
...
176530 TTGGCGAACAGCCTGCCCGCAATATTTGCCGACGGTAAATTTGACCGCTCGCGAGT |-----|

165029 TGGCTTTCGCCACACAGAGCCATTGACCGACTGCTGCAACGGCTAGCCTTAAGCGCTAA 165019
165089 GCGTAAGTTTTCGCTGTTTGGCACTATTGGAATTCAGTCTTTACGGGAATCTGAGACCCC
165149 GGTATGCCCGCATCTGCTTCGCAACCCCGTCGAAACCAAGGTGCGCCCCAGAAATGGTT
165209 TGCAAAATTATACGGATATTGTGCGGTGCTGCCAAGTC
165259 TGTCACTCTTCCTGCCCTTAATTTCGGTTTGAGCAGGATGCGGACGCGCGTGGTGCCT
165319 TCCCGGGGTTTCGGGTAGCGGAGTACATCGGGGTGTTGCATCAGCC
165478 AGGACTTTTGGGCTTCTTCTGTGTGTAGCCGTCAGGTGACATCGGTAACGACGGGA
165538 TAGTATCCGTTTTCAGCGGTGGATGCTGTTTTTTCCTGTCGCTGCTGAAAGCTG
165598 GCGGGCGGTGCTGTGTATGCTGCCGATGTAGAAGTAGTTTCTTCTGTCGCGCGGTTG
166011 TCGGCAGTGTGATTGCCGGAATGCGGGCGGTCGAAACCGATATGCCCGAAGCCCTGAT
166071 AAAGTTTAAAAACCTGCCTTGCGAAGCAGGTTAATGGTTTGCCA
166349 CGCAGCGGCGCACTTAACCTGTTGGCGGCTTTCAACAGGTTCAAAACACATCGCCTTCA
166409 GATGGCTTTGTGCACCTACCTTTAATCAGTCCGGAATAGGCTGCCCG
166619 ACCCAAATACCCCTTAATCCTCCCTGGATACCCGATAATCAGGCATCGGTGCGCTTTT
166679 GCGCGCAGCGGCGCACTTAGCCTGTTGGCGGCTTTCAACAGGTTCAAAACACATCGCCA
166739 CGGGCTTCGACAAATACCGGTGCGGCTTTGGTTTGGCTTCCGTCAGCG
167000 TGTGTGTTGTAACCGAGTTTGTAGAGCCGTTTTTCTTGATCCAACGGGATCGCTGTC
167050

167060 TTACTCGGTGTGTTTTGGCCGCTGATTTGTCTTCTCATCGACTTCTATGCGCTGAGCG
167120 TGTATTGCTGCCGGCGTCTGAATAATGGTGGCATCAAGCAGCGCGCGGATGCTTTCTCT
167180 ACTTTTAAACCTTTTTCGGTCAGTTGGCGGTTAATCAGTTCAGCAATTCGGACAGGGTG
167240 TCGTCTTGCAGCGCGGTTGCGGTAGCGCATAGGTGCTGTATATCGGGATGC
167259 AGCCTTTTCGACATAGACGGTCTGCTTACCTTTGGGCAGTCTTCCAGCAAGCGCACAGT
167819 GTTTGCACATCAGGGCAATTCGGCGGGTGTATGTGCAGTTTCTCGAT
167955 CGGTGCTGTTGGCGGTTGATTTGTCTTCTTCATCGACTTCTATAGCCTGACGCTGTTT
168015 GCTCCGCGGCTCTGAATAATGGTGGCTCAATGCTTCTCTATTTTAAACCTTTTTCG
168075 GTCAAGTTGGCGGTTAATCAGTTCAGCAATTCGGACAGGGTGTGCTCTTGGCCAGCCAG
168135 TTGCGGTAGCGGCATATAAGTATAGTGGATTAAACAAATCAGGACAGCGCAAGCCG
168195 CAGACATACAAATAGTACGGAACCGATTACATTGGTGTTCAGCACCTTAGAGAAATCGT
168255 TCTCTTTGAGCTAAGCGGAGGCAACGCGTACTGTTTTTGTATTAATCCACTATACTGTC
168315 AGTTCGTGGAACGGCAA
169111 CAAAGGTATTCGAGATAGCGCGGGCGGTTTTTTGCGTGCTTTTTCGGACACGCGTG
169171 TTGACGACGCGGAGCGGCTTCGGCGGAATGGTGTAGCTCGGATAGAGATTTCAAAC
169231 TGTCTTGGGTCAAGTTTTTGTGACGTAGTTGCTTGTGCTTCAAAAGTATGATGAC
169291 TCGCCGATGTTGCGTTGTGTGAAGTGGTGGTGGCGCGGCTCCGCGTTTTCAA
169295

...
161689 TGAAGTGATTACGCCCTCTGATGTAAGCTGCTAAACTGCTAAATCTGTTAGATAGGACGATACGA
161749 TGAAGCCAACCTGTTACTCCCTGACTCGGCATTTACAGCAATCAGT
...
161985 TGCCATTATGACATTTGTTAGTGATATGAGGTTTGATGAGATGCAAAATTTGGCGACAAAT
162045 TCAAAATTCGTGAGGCGTACAGAACAACGGTATATAAAACATTCC
...
162302 AGAAACCCAAAATCTAGAACGAGTTTTTAACCCCTTTTAAAAATCCCATTCAGACGACCTTT
162362 TCTAAAATCCCTGATTTGATTTTAAATTCATACAGGATTTTTC
...
162543 GGCTCCATGGTGCCCATGACGCGTGAACAGGCNAATCAGCATTTACCGTAAAGCGTTTGG
162603 GAGCGTTTACCGCGCGCGCAACCAATGCGGAAGCGGTGCGGTTCCTCAATTTTGTGATGCC
...
162670 ACGGTTACGGCAATGCCGCCGTATGCTGCAACGCCGCCGAGCGGTACCGGACGCGGG
- |
162730 TTATCGGAGCAGTCAACCTCAAGCCATCAATTCCTTCCCGAAACAGACCTTTATTGC
162790 GGTTCACCGCGGAGCGCTCTGGTCTTTTATACCAAGCTCGGTACGTTACCTCTCTTCGGCA
162850 AGGCTGG
...
162893 TGCAGATAACACTGATTAAAGGAGATAAACCATGTCAAAAAGTCACTCATCGCCCTAA
162953 TGACCGCAGCCATGCAGCCCGGATTTTCAGCCACAGCGACCTTAGGCATTTGCTACGCCATGC
163013 CGACTCAGGATGTGGACGCAAGCCCAACGGAAGCGGGTAGCCGCGGAAACGGG
163073 CAGCCAAAAAAGCGCTCAAAAATAACCGCCTTTTCCGATGGCTGGCGGCTTGCTGCTC
...
163133 TAATCGGCCACAGGAAAAAATCAGCCATACCAAACTATGGCAAAAGCTCGCGCGACGCCG

163193 CATGACTTGGAACTTCGTGCAGCGCGGGACGCGCCGGAATG
...
163245 GCTTATGGCCATTTGGTCGGCGGTATGCATTAAATCAAAACCGGCATCGCGCGCATTCG
163305 CAGTTGGCAGAAATCAAAAAATCCGCAAAATCAGGAAGGGGGGCGCA
...
163843 GGCAGAAGACAAATATGCCGATTTGAACGGCATATCCATGATATTATCGACGACATAA
163903 GGSAAAGATACGGCTACCGGAGGATTGCGGCAGCCATCCGTCACGCA
...
164026 ATACCGCTCGTTCAAAGGAGAGTGCGCCAAAATTGCGCCGAATATCCTGCGACGCTGTTT
164086 CCATGCAGAAAAACCGCAATGAGAAATGGGTAAACGGACGTTTACCGAGTTCAATGTAGCGG
164146 AGAAAAGATATACCTTTTCCTCGATTATGGATTGTTTAAACGGGGAATCGTCAAGTTACCG
164206 TATTCAAACCCCGCCGACTTTTCGATTGCGCGCGAGATACTGAAAAGGTGCGCGCGGAGAA
164266 ACGGAGATCGTCTGAAAAAGCCGATCTGCAATTCGGATCAAGGTTGGCAATATCAGATGTT
164326 TTTTATAGTGGATTAAATTTAAATCAGGACAGGCGGATGAAGCCGCGACAGACAGTACAAAT
164386 AGTACGGCAAGCGAGGCAACGCCGTACTGGTTTAAATTTAATCCACTATACAAACACGA
164446 TAGAATCAAGTTGAAATTTAAAGGGCTGAGCCTTGTTTCAGTACAGAATTCA
...
164624 GTTTTGGTTAACCTTTTAAAAAATGCGTGATGACTTTTGCAATCTTAAAGCGCTTTT
164684 GGGTAATTCTGTGAAAGAGTTACCCCAAAAGTTACCCCAATAAATG
...
164833 GTATGTTGAGGCGGGGGAATCGAACCOCCTCGGAAAGTCTCTACAAAGCGTTCTAC
164893 ATACTTAGTTGTCTATTGTTGAAAAATCTTATTTCCATCATGCGGAC
...
164969 AATCTCATTTCCTGCGCAAAATAACCCGCGAGGAAACCAAGTCAATGTAGATGACGTTGCGG

```
157347 GTAACAGCCGGTGGCCCATCATGGATACCATCGGATCTTGGCTATTTTACG 157337
      |-----|
157407 GATTCGGGTTGATTTTGTGCGCTCTGCAATGATTTTGGCGATTCGACATCTACTTTT 157412
      |-----|
157467 CTCAGACTGGAGTGTTTCATTTCCTGTTGTTGATTA 157445
...
157827 TCCGTGCATTTGTGTTTGGATTTCGCGGGTGCTGCCGTATAGCCAGTCGTCGGAT 157877
      |-----|
157887 GCGGTGTGGCGGAATTTGATGCTGATGTGTCGGGATGGGGGGGGTTTGGCGCGTGG 157887
      |-----|
157947 GCGGATCGCGGGGAAGGTGATGTTGTGCGCGGTCTGTGTGTCAGCTCTCGGAGTAT 157974
      |-----|
158007 TTCTACGCCCGAGTCCGACAGTG 157974
...
158196 GCAITTCACCGCTTGCGCGCACCGACATCAACCGCTTGCCGTGCGCTCATCGCTCCA 158246
      |-----|
158256 CCAAAACGCCGAATCAATTTGGCTTCCTCCGTTTAAGCTCCGCTCTTGCCCGCATCGC 158311
      |-----|
158316 TCGCGCGTTTGGCGCTCGGCTTGACGCTGACCGCTCCAGCTTGCGGTATAGCGTGGCA 158338
      |-----|
158376 GGCTGATGCCAATTCCT 158338
...
158416 AGACGTCGCGCGTTCCATTGCTTCGCGCTGATTCCTGACTGCTGCTTAAGACGCTCAAT 158466
      |-----|
158476 CATTGCGGATTCATCGCCTTCTCCCGTTTCACGCCCAACCAATTCGCGCACATTGCTG 158489
      |-----|
158536 TCGGTGCTT 158489
...
158677 TCAGACGGCTCACTTGACTGGGATACCGACCTCAAGGCTTGTTAACTGCATCGCCAACT 158727
      |-----|
158737 CGCTGCCCTAGCTCTTCGCGCTTCGGCTCTCTGACAACGGTTTGCTT 158727
...
158826 GTTTTCATCACTTCGTCTTTGGCGGAGGTTTTCGCGGCTTTCGCGCAGGGCGACGCGC 158876
      |-----|
158886 AGCTCGCGCACCGTCAATTCGGTCCACATCGTCAAGGTCATGCGGTTGACTTCTTCCCT 158912
      |-----|
      |-----|
```

```
158946 TCGGCCAAACCCACCAGCGTAACGTCCTTCTTCGACCAGCAGCTCAAGCAGCTTCGACTTG 158966
159006 CCCAAATCCATCAGCT
...
159145 TATGCTCTTTTAAAAATAATCAGCGCAGCTCCAGTTTCGAACATGCCCTTCGCGCTTCCGCTGCC 159195
      |-----|
159205 GTACTGCTTGACGACCGCGTTCACCCCAACGCTCTTCGCTATAAAGTCTCGCGCTTACCCC 159232
      |-----|
159265 ACTGCTCCATCACCAGCACACATGCGCATGCGCTGCGCTGATTGCTTACTTTGTGCGTTGCCA 159291
      |-----|
159325 TAAATTCAAATTTCTGTATTCATTTTTTATCTCCAAAGTTTCGAGCTCGGAAACTTTCAA 159367
      |-----|
159385 AATCGTTTAATCCACATCGACCGCTTGCGGATTTTCGG
...
159519 TGACCAATCCCGAGGCAACCAATCATCAATATCCCTGCTGACTTGGGATGCGGTCAGCC 159569
      |-----|
159579 CCAGTCCGACCGATAAATCCCTTATTGCTCAGACCGATAATCGGATG
...
159764 CCAGATACACATCGCGCGCTTAAAGCGTTCTCTCTTGGCCACACTGCGCCAGCGTTGGC 159814
      |-----|
159824 CGTTTGGGCAAAATTTTCTTTTAGTTTTCAAACAGTATAGCCATA
...
159961 AAATCTCTTTTGCAAGGATTTTGTGATTTCTAAATCTCTTTTGGTAATCGTTTGA 160011
      |-----|
160021 AGAAAAAGAAAGTTTTTAGGGCTTACCCAAAGCACAAGCTGCTGAA
...
160215 CTGAAATTAGTAAGGAGAGTCAGAGTTATTAGGGTATTACCGCCAGCCAGCGTAACG 160265
      |-----|
160275 GCAATTTGTCTATTCTCAGTGTGGCAAAAGCGCAGAGAAAAAGC
...
160390 TTATGTTTGACTTTGAACAACAAATCAAAATGGGGTGAGCGAGCGGAAGAAATCGTTAAAG 160440
      |-----|
160450 AGGCAGCCACCCAGACAATATCGAAATTCGCCAACCCTTTGCGCTTC
...
161558 AATTTAGAAGTCTTGTAAACGGTAATTCGAGAGCCAGTATTGAAGCGCAATCGCTGATG 161608
      |-----|
161618 CAGGAGCAGCAGGAAATTTATTGATGATGAACGGTTAATCAAG
```

153637 AACCGCGCGCGACAAATGCGCGCTTACTTCGTGAATGCCTTAT 153627
...
153926 TTTTAAATTGCTATCCCATGCTTTTCGGTGATTTGTAGATGTTGGTTTTTGGCGTTACAA 153976 153984
---| |-----| |-----| |-----| |-----|
153986 CGACCAGCGCCCACTGTGGTTTGTATATTCAAATTAATAAGACAGTAGTGCAT 154044
153989
154046 CGTCAAAATCGGGCGTAATCAGACAAAACGGTTCGCAGATACCGCTTAATATTCG
154049
...
154178 TTTCCCGTAAGACACCCCATAGGTAAATCGCGCATTAATCTAAATAATCAACGATTTT 154228
154238 GAGTCAATGCGGCGAGTAGGCATTGCTGAACACCACGCTTCAAAAAA
...
154308 AACCATTTGGAGCAATCAGCGGTTGCCGACTTGTGCGGACACCAGATGCGTTGGTA 154358
154368 TCTTTTCCACTATCTGCGCTTTCACTATTGCGCCTTTCAGGCTG
...
154569 ATCTGTTTGAGTGATAGCAAAATGGTGGCTGCCGTACAATCTTTGGCGATTTC 154619
154629 TGCAGATAGCATCCTGGTGTGCCCAATATATTGAGCGGTTTTT
...
154729 TTTGTTTTTTAAGCGGAATCCACAGGTAAAGCGTGTCTTGACAACTTAACGTTGCTG 154779
154789 CGGTTTGGCTGATGTTTTTGCATTGTTTCGTAATATAGTGGATTAAAC
...
154842 AGTACGGTGTGCTCGCCTTGCCGTACTATTGTACTGTCGTGGGCTTCGTCGCCCTGT 154892
154902 CCTGATTTAAATTAATCCACTATAGTTTAAAGCTTTGTTCTTAA
...
155030 TCCCTTCGTTGGCTTTGGACTTTGGCATACCGCATATCCGAGAAAGGAGCGATGATT 155080
155090 TACTATCCCCTGTGCTGTTTCCAAGACCCCGATTAAAGCAACTGAGGCGTGATTATCCC
155150 GATCTCGGGCTGCTATCCTGTGCGGATACGGTTGCCGACGCGTTGCCGATACCAA 155174
155210 GCCCGCGCAATGTTCCATCT
...

155272 AAGCATAAAGGTTTCATCGGACAATCCCGATTATGCCTAGCGGTGTTGTGGACGATGTT 155322
|-----|
155332 GAAGCGGACGATGCGGCATTAAACCGGACAGGTACGGTTTAATATCATGTTGGCCGCAACAT 155338
|-----|
155392 AT
...
155417 CTGCCGCGTCATGAACCCGAGCGGTTTTTTGGCGAAAGCGCTTTAAAGCGGATGCAG 155467
|-----|
155477 CCTTAACTTTCAGCCCGGATTTTTTGGGCTTTTTTATTATGCGCGATGC
...
155910 GCATCCCGGATTACAGCACCTTATGCGCCTACCGCAACCGGCTGGCGCAAGACGACGCC 155960
|-----|
155970 TGTCCGAATTGCTGGAACCTGATTAAACCGCCAAC TGACCGAAAAAGGTTTTAAAGTAGAGA
|-----|
156030 AAGCATCGCGCGCTGCTTGTGATGCCACCATTTTTCAGACCGCGGCGGCAACAGCGTC 156044 156084
|-----|
156090 AGGCCATAGAGTCGATAACAAAGGACAAGTCAGCGGACAACACTACCCGAGTAAGAGC 156149
|-----|
156150 GCGATGCGCGCTTGGACAAGAAAAAGGCGCTCTACAGACTCGGTTTACAAACAAC 156154
|-----|
...
156395 CGGCAGCCTATTTTCGGACTGATTAAAGTGAGTGCGCAAGCCACCTGAAGCGCGATGTGT 156445
|-----|
156455 TTGAACCTGTTGAAAGCGGCCAACAGCGCTAAGTGGCGCCGCTGCCG
...
156726 AAAAAATGCCGTACCCGCGGCGAGGCTTCGCACCTCTGAAGGTTTCAGACGCGCTTGT 156776
|-----|
156786 TTTTCTGTCCAATCGAGGAAC TGCCGCAATTTTTTCCAGCGGCATAT
...
156849 GTATCGGGTTCGGCTTCCCAGCGGCCTACCTGTATGTAGTGCTTCAACCCGACGATTTGC 156899
|-----|
156909 GCCAACTCGGCCTGTATCAGTTTGCAGGGTTGCGCAGGTTGTAAGGTGTGTAGCCGAGT
156969 TCCATGTCGTTGCGGTTTTATTTTGTTCGCATATTTTTTTCAGTGCCTCCGCGGCGAGGTTT 156974
|-----|
157029 C
...
157287 CCTCGCGGGCTTGTGTCAGGCAAAATTTGAATATCAGTGCCAAACACGCGCGATGCG
|-----|

150506
150516 ATCGAATGGCGAACAACCCCTCAACTACCTGACATCAAAACCGCTCGTTCAGTTGGGCA
150557
150560
150576 CTTGTTACGACCGCTCGAAAATACCGCTTTGAAGAAT
...
150690 GTCAGCAATACCTCTACCACGATCCTTGCCACACGCCCATCAAAACCATGAACGCCACCC
150740
150750 AAATGGCAGCAGCGCTGATGGGGCAGAAAAGTCTGCTTAAGGACCGCTGCTGCGCGAAT
150779 150788
150810 CTGGCATGTTCCCGCTCAACACGACCCGATATCGGCCACCCAAAGTCAAAATTCGCGCAAAACAAG
150828 150839
150870 AGGAAATCGAGAAAAACCTCAAAGA
...
150918 AAATGCTTACTTCTGCTGCCAGCCTGCCTGCAAGGCTTGAGCGGCTTACACCGGACGACAACA
150968
150978 ATATGCTGCCGACTACATCTGCTCATCGTAAATGGCGAACAACATCTCTCGGCGAAAACACTGGC
150995
151038 TAGACGAGTTTGT
...
151068 CGGGGTGGAAAAAGTATTGTGTAAACAACGGACACGGAATGCCGTCTGAACGCCGAAA
151118
151128 GCCTTCAGACGGCATGTTTGAATCAAAATACCTACCCGAATCATGA
...
151312 TTACCCATACCGTGCATTCGCGGCCACATGGTTAAATACGTCCCCGAAATCGCCCAAG
151362
151372 GAAACAAAACGTTTTTCGGGACGCGCTGCACCGTCCAACCTCTGGGCGCGCATCGGGACA
151379
151432 ATATGGCGGGAATGCGCTGGAACCGCTCCGCTTCGGTGCAGA
...
151519 GGCACAAAGGCGGCGCAATTCTTCTGAAAGAACGGAACTGATATTCACATCGTCAAAA
151569 151577
151579 CGCTGCGGGAACGTTTTGCCCGCACATATCCGCTCAACGCAAAAATGCGGCTCGGCTATG
151582 151593
151639 AAGAC
...

151726 AGGGTTACGAACCGCGCGCATTTGGGAATGATAAGGAAAATCCGAGACACGCTCGATA
151776
151786 TCCCAGTTACCGCAACGCGGATGCTTTCAGCCTGCAGACATATATCGGCATCAAAACA
151803
151846 TCAGCGGCTGCAACAGCGTCTCATGCTGGGGCGCGGCTCATCGCCCGGATTGGGCG
151860
151906 GGCAATCAA
...
151978 GGCAGTTTTTCGAGCTGTGCCTGACAAAAGAGGCAACAATAAATACCGATCGCCGCTC
152028
152038 TTAAGCAATGTTGGGCATGATGAAAAAACGTTTCGACCCGGCGCAACGCTGTTTGACC
152066 152084
152098 GCGTCCGAACGTTTAAGGATCGGATGAAGTCCGCCGAATCC
...
152233 TTCCCTGATTTTCGGCGTATTTTGTGTTTTCAGTTTCGCCCCCAACCAACGCTTTC
152283
152293 ATAATCCGATTGCGGGTATCGTCTTCTGTCATACCGAACGCGCGGATTTGACCCACAG
152311
152353 CGACAGCGCGGACGCAAAAGCGCCAAAGCCGATTACATACCAAAACATTTGCCCTCC
152356 152365
152413 CGATTTGT
...
152517 AGTACAAATAGTACGGAACCGATTCACTTGGTGTTCAGCACCTTAGAGAATCGTTCTCT
152567
152577 TTGAGCTAAGCGGAGCCCAACGCTGATCGGTTTAAATTTAATTCACTATAAATCTTCTGA
152594
152637 CCGTAGTATTCA
...
152859 TATCAGGATGTCATGATAGATAGCATGAATTTTCCCTTCTCTCGCCCATCGTTGG
152909
152919 TCACAGCTCATACTGCAGCCAATCGAATTCGTGGTCTGCTACCCCTGC
...
153185 CTTGTTTTTGGAGATGTTCCGCTTAATTTGCGCCGTAACCTGTCGGCGGAGCGTCCCTG
153235
153245 CCGGTATCGGAATCCCTCTGTATGCGCGGAGGTGTCGCGCGTATCC
...
153577 CGGCACCGCTTCTTTTCATGTGCGGGCGGTGTACGGGCGCAGCCGGACGCGGACGCTGCC

147582 TGTTGCCGAGCAAAAGCAGTCGAGCGTTTGTATAGCAGTTTGTGATCAAAACCAGCC
147549 147557
147642 AAAACGCCATCAGATTTTGAAGCGCTTAAGCAAAATTAACAAACCGCACAATATCATGT
147663
147702 TTGACGGATTGGCCCGGTAAACGACGTTACTGATGCGACGAGCTGGCGGTGAGTATC
147767
147770
147822 TCGA
...
148052 TTGCCGTCGATCCGACTTCGCGGATGCGTCTCTGCTGGCGGTAATGTGGCGATGAAC
148102
148112 GCAGGCGGTAAAGCCGTTGTGGGGACGCGTTGGACAACTCGCGCTACTGGCAG
148144
148172 ATGCTCAACCGCAGGCGGAATGTTGC
...
148209 GTGTGCCCAACAATTCGGCAAAATCCACGATGAAGAACCAGCGGTTCGACGTTTCA
148259
148269 CGTGGATTAGACGGGATCAACATCGTTAAGACCGAACGCTTGGAAATCCCGGGCCACA
148292
148329 AATTCCGCAAGTCGGTTTGGCAAGACAGCTTACCGACAAATCTTGGAGCGGCTGCCCG
148355
148389 GCGTGCAAAAGAGGTACAGACGGCATCATCACCGGCTCGCCTTCGTGTGCAAAA
148424
148427
148449 TGCCGAATACACGCGCACGCTGTGTATGGAGTTTTCGGCACGCTCGCCACCGCCACGC
148493
148509 CTTCTATCGTTGAATCCGCGATTCTCTGTCGCCCCACGACAGCGTTTCGGCTGGCAGTT
148514
148569 TGGNACATTAGACTGGCGTTACGTCGCGCGCTAGCGCCACCAAGCGCGGCGCA
148584
148629 AGGACACCGCAAAATGGTCTTGTGTCGACAGCGTGGTTTCAGACGACGAGCAAGCCCGCTAG
148632
148689 AGGCAGCCCGCAGCACATCTGTGAATTTGCCCGCGCCGCGAGCGGAAGGCTTTATCG
148745
-| |----| |----|

148749 CCTATGCCCGCAAGCGCAAAACCTTCTGGCTGACCGCAGCGCACCGCCGCTCG
148750 148758
148809 CCAAGCACCAACGCGCTTTAAATTAACGAAGAGTGGTCATCCCACTCGAAAGGCTCG
148810 148820
148869 GCGAGTA
...
149271 TCCAAGCTTGGGAAAAATCCACGCAAAAACCGTACGAGCGCGGTCTTTGTGCGCCTGC
149321
149331 ATATGCACGGCGGACGGTAACGTTTACACCAATATTCGGTTAACTACAGCATGCCG
149342 149351
149391 AAATGCTTCAGACGGCATACCGCTCAGTCGAACCATATGAAATCGCCGCTTCGCTCG
149435
149451 GAGCGGTGATTCGGGGAACACACGGCATCGGCATCACCAAGCTCGAATTTTAAAGCGACG
149504
149511 AAGAAATGCAGCCGCTTTTGGACTACAAAACCAAGTCGATCCCAACACACCTTCAACC
149571 GTCACAAACTGATGAAAGGCTCGGACTTACGCAACGCTACACCGCTTTCGAGCTTT
149603
149631 TGGCGCGGAATCGCTGATTATGAAAAATCAGACCTCGGCACGATTGCCGATTCCGTC
149641
149691 AAGACT
...
149784 GCGTGGGTTTGTGACCGAGGCTTCTATACGAAGAACAACCGCTGCGCGGCTTTCCA
149834
149844 TCAAACTTTGAAGAACTCATGGACATCGCGACCACTGCACCGTGTGCCACCGCTGCG
149873
149904 TCAAACTTGGCCGTCACATCGACTTCGGCGACGTAACCGTAGCCGTCCGCAACTACC
149933
149964 TTGCCGATTCCGGCCACAAACGATTTGCGCTGCGCAGCTATGGGTATGCGGTTTGA
149982
150024 ACGCCACCGGCCGAAAAACCATCAAGCCCTTCGCGCGCCATGATACAAACCGG
150028
...
150301 CGCGCCGAAGATGCCGAAGCGGTCTTCTACTTCCCGGTTGCGGTTGGAACGCTGTT
150351
150361 CAGCAAAATCGGACTTGGCGTTCAAGCCATGCTCTGGCACGTCGCG
...
150456 ACGCGGGCGCAACAGGCAAAAGCCGAAGAATGAGCAACCAACACCGCTCGCCTTCC
150455

144491 TGTGCCCCCTTTGCGCCGAACAGACGGAACCGCCCAAGCCCAATACGTTAAACGCCGAGT |-----|
144541
144551 CTGTGGCGGGCGGTGCGCCGTAACAGCTTAACCCAGTTGGGCAACACAGCCAGCACGAATATCGCGC |-----|
144564
144611 CCGTCAGCC
...
144780 TTTGGGAAGTAGGATCAGACACTTATAGTGGATTAAATTTAAACAGTACGCGCTTACCT |-----|
144830
144840 CGCCTTGCCGTACTATCTGTACTGTCTGCGGCTTCGTTGCCTTGTC
...
144900 ATCCACTATATTTGGGAGGGCGCGCCGTTGTGCGGATACGGTTTGAAGCGGATTACC |-----|
144950
144960 CGACGGGGAACCTCAAAACCGGACAATGCCGTCGAAACGGGCTTTTTCGCTTCAGACGGCAT |-----|
144997
145020 TGCCTGCTTCAAGCGGACGCGCTTATTCGCG
...
145111 AGTTTGGGTAAACTGTTTGAGCGCGGTTTGGCTTCATACAAATCAGCAGCTTGTGT |-----|
145161
145171 TCGCGGGGATATTGCGCGCGCGCTTGGTCTAAAGATGCAGCATTTTCGTGTTCACAGCC |-----|
145225
145231 ATGCAGTCGTTGGCTTTTGTGATGATGATATCCTGTATGTAAGTGAACGCTGGCACATGAATGT |-----|
145236
145291 C
...
145306 TGGATGATTTTATTTGTCGATTTCCGACGGCAGTTTTTCAAAACAGTCGCGGATCGTCGCG |-----|
145356
145366 CGCAGGGGATGTCGAGCGCGCAAGTCTTCCACTTCGTATCATGCACGGCGGTAG |-----|
145371
145426 CGGATGGCGCGGAGCGGAGCGGCAACGGTGCAGCATCCCGCTTGGGTTTCTTCGCGG |-----|
145440
145486 -| TTGCACTCGAANAACGCGCGCTGTGGTGTGCAAAATATTTCTTTTAAGAACG |-----|
145487
...
145849 CTTTCAAATTCGCCAAGCGCGCGGATGTCGGTCAGCTCGTTGATGTCGTTCTCGT |-----|
145899
145909 CCGATGTAGAACGCGGGTTTCTTTTCTTCGGGGAAGGTGCCA |-----|

146113 TTGTTTTGATAATGATGACCTTGCAGGTTGGTACGATTCTTCGGGGTGTTCGCCCAAG |-----|
146163
146173 TCTATGCCCAAGTGGTTGACCAAGTTCCAACTGCCTTCTTCGTTGA
...
146353 AAAGTTTTCAGCAGCAATTCAGCTGGTTCAGGGTCGAGCCGGCAGCAGCAGCTTGT |-----|
146403
146413 TCGTTGTTGATGATGTTGGATGATGTTTCATCCGCATGGTGTTCACAATCAGCATGTCG
146473 CGGTCGTAATCGTTGCGCGTCGCGGTCGAGCCGCTCAAAACGGTATTTCGCCGCCCTGC |-----|
146481
146533 GTAA
...
146564 CCTGCAGAAATTTCCACATTTCCAGAAATGGTTCGGGGCGCACCCGCCAACGCCCTGC |-----|
146614
146624 CCTCGCGAAGCGGTAGCCTTGGCGGTATTGTTGGTCTTCGCGG
...
146817 AAAGAAATGCCGTCCTGAACGGCTTTCAGACGGCATCGCGCGGTACGAGAACGCCCGGAA |-----|
146867
146877 CATCGAAGCTCCGCGCTTCAAATAGGGCGGCGCAACCCCGGCGACTGGCGCACCG |-----|
146889
146937 GAGTGGGC
...
147065 GAAACACCACCTAAATTTATATTTGATAATATTTGTCAACAATTTCTCAAGCGTTATTT |-----|
147115
147125 TGTTCCTATAAGGTATTTCCTGTTTCGGCATTGAAAAATATCAAAATTCACACTACATTT |-----|
147185
147185 ATCGCCCTTTCAAACCCGCTGAAACCGACTTTTCAGACGGCATTCAAATAAAAT |-----|
...
147282 AGCGTATTCGGGAAATTCCTATTAATATTTGATAATATTTGTCAACAATTTCTCAAGCGTTATTT |-----|
147332
147342 GATTATTGGCGGAGAGCGTGGCAATCTCTCAAGACCTGCGGGGACAGCGTAAACCGG |-----|
147350
147402 GCGCTTCAGCACGAATGCTGTTTGAAGTATTGGCGGATATTTGGTGGTGTGTCGCAATC |-----|
147402
147462 CGTATCTGTGTCGATGACTTGTGGAACATCCGAAACGCCCGCGGCTGTTACGGGAAA |-----|
147495
147522 TCGCCACAGGTTGAACGAATCCGCAACGCCCGGACGATATTCGGAAGTGGATGTCT |-----|

139638 TTGAGGTAACACTACTATGATGGTCTACAAATTCCTTCTGCTGTAGTCATTTGATAAT
|-----|
139679

139698 GATATCGTTTCCAATTCGCGCTGCTTCCATFACTGGCTTGAATGCTCCTACGACAGAG
|-----| |-----|
139701 139756

139758 ---|CGTTACGGATTCTAACTCTCTTCTTCATCTTTCATCGGTAGCATTCGCCGTGAACCGTT
139761

139818 GCAACAACAGTGTGCGCAATAGCGGTCTTTCACGGTTGCGGAGCGGTTTGGTGTGG
|-----|
139830

139878 TTGCGTGTGAGTCGATACGCGACCCAGGCAATTGAGGGCACTATTCCAAATGATGCGG
|-----|
139898

139938 TATATTTTGTTCATTT

...

141319 TTGCTTGTGCTGATTTCCCGCTGCGGGAATGACGGCGCGGACGCAAAACGACCATTA
|-----|
141369

141379 CACAATTAATGACAACCCCATTTATTGCGAAGTCAGCGCTAGGAGA

...

141573 CACGCTTGGCGGGCGCAACGGATATAGTGGATTAAACAAAACAGTACAGCGTTGGCT
|-----|
141623

141633 CGCCTTAGCTCAAGAGAACGATTCTCTAAAGGTGCTCAAGCACCAAGTGAATCGGTTCCG
|-----|
141650

141693 TACTATTTGTACTGTCTGCGGCTTCGTCGCCCTTGCTCTGATTTTGTAAATCCACTATAA
|-----|
141718

141753 AACGTGTTCCCTGCGCGGAGGCGGAACGCCGATGACGGGTTTTCCTCAAGGGTGC
|-----|
141780

141813 GGCTGCCGTATATCACGAAATCCAACAGGTAGAAATCTTCTTGCCACGCGCATTCG
|-----|
141884

141873 GGGCATTTCCAGTCTGCGGGGATGCTTCAAACTTTGTCCCGGGCGGATGCGGTGTTG
|-----|
141884

141933 GGGTCGCGTGTCTTCTCATCGTAAATCCAGCGCAGGGGCCGACATATATTGGCCCAT
|-----|
141935

141993 TGTGTTTCCTGTTTTTTGTATAGTGGTTAACAAAAACGGTAGCGGCTTGCCTCGCCT
|-----|
142038

142053 TAGCTCGAAGAGAACGATTCTCTAAAGTACTGAAGCACCGGTACTATTGTACTGTCTGC
|-----|
142065

142113 GGCTTCGC

...

142462 TCGATTGCCGCTATTCTTTGCGGAGGTGTTTGTATCGGGGAGTTACGATGGCGACAAC
|-----|
142512

142522 ATTGCTTCGGGACGCGCTTTGGCGGGACTCCGCATCAGGTAGC

...

142693 TCGGTTCTGCCACGCGCGCTCAGCCTTGTTTTGAGGCTGTCGTAGCCGTGCTCGAGA
|-----|
142743

142753 TAGCGTTGACTTTCGGCGTTGACAGCTTGGCGAGCGGATGATGCCGAGGCTGCCGTCT
|-----| |-----|
142767 142776 142773

142813 ATCACCCCGCGCGCATGC

...

142846 ATAAAGCCTGCTTTGATGCTTTGGATGTACTCGGAACCTGTTCCGGCGCGGATGATG
|-----|
142896

142906 TCTTCGTCGGGATAAATACGTCTTTTGAAGTTTCAGGCTGAAGGTG

...

143207 CTTTGGCTTGCAGGATGTCGGCAAGGATATTGTCTTTGACGGCCCGTTTGGGCGTTGGT
|-----|
143257

143267 GCAGATACCACGCGCAAGCGCTGACACCACTGAAAGAGATTGCGCCGATCTTT

143327 TGCCGATTTTCGGCAATACGGGATTTGCGTTGCCAAACCCAGCGTTTCCCGCTTCGG
|-----|
143346

143387 CTTCTGTACCGACTG

...

143419 GCGCCGAGTTCGCGCATAAATGCTTCGGGTAGTATCCTTTGCGGTGATGCTCCACT
|-----|
143469

143479 ATGGGTTTGAGCTTGTTTTTGACGAATTCGGCAACGTTGGCAATCAGGGTTTGGCGGTTT
|-----|
143511 143532

143539 ATCTTTGTTCTTTAAGGTTTTCGGGGAATTCGGGCGCGCTGATCG

...

143838 TTTTTTGCCATTACGGCTTGGTTAAAGCGCGGAGCGCGGAGTTTAGCATCGTCAAAAT
|-----|
143888

143898 TCGCCTGCGCCAAAGCGCGCGGCTTTGACGGATTGTCGGA

...

144009 GCGGCAATAATCTGGATTACTTCGACTTCGCCGCGGTGAGGCTGTTGGCGGCTGAGC
|-----|
144059

144069 TTGCCGACTTCTTGTTAAACGCCAAGGCTTCGGGGCGGTTTGATAATACGCCGATAAG
|-----| |-----|
144087 144104

144129 TTGGGGATAAAGCGGTTGTTTGAAGTACCG

...

```
...
136291 AAATCCCGAGACACTGGCGCTGGGTGAACAGTGGGATAGGAAAAAGCGACGACT |-----|
136341
136344
136351 GGGATAGATGCGCGCGCCCTACACTTTGGCGCGCGGGGTATTGAGGA
...
136401 TTGAGCCCTGGGANTGTGGCGGTTTATGTGACGTACAAACAGCAGCATGACGATGCTG |-----|
136451
136461 CAGGATTCGAGTACCGAAGATTGGCCGCGCCGACCGTGGCAAGC
...
136521 GCCGATGGCTTTACGAAACCGTATCCGCCAATGGCGCTGTGATGCCGGAACGTCAAAA |-----|
136571
136581 |-----| |-----|
136583 CTGGCAGAGGCTTTGGAATTGATTGATTCTTGATGGGCTTTGGTCAGAAAAAACACCC |-----|
136616
136641 AAACATTTGGCTGGCTTTGTGGAGGTATTGG
...
137038 CGGCGAGCGGGCGCACCTTACCTGTGGCGGCTTTCACAGGTTCAACACATCGCCTTCA |-----|
137088
137098 GGTGGCTTTGGCACTCACTTAAATCAGCCCGAAATAGGCTGCCG
...
137326 AGCCTTTGTGGGATAGACGGTGTGACCTTTGGGAGACCTTCCACAGACGCGACAGT |-----|
137376
137386 GTTGCACATAGGGCATTTGGCGGGGTGATGTGACGTTTCTCGAT
...
137457 TGTGTATTGAACGAGTTGTAGAGCGCTTTTCTTGATCCAGGGGATCGCTGCC |-----|
137507
137517 TTACTCGGTGTGTTGGCGGTGATTGTTCCTTCTTCGCAACTCTATAGGCTCGACGC |-----|
137572
137577 |-----|
137577 TGTTCGCTGCGGGGTCTGTGATATATGTGGCATCAACGAGCGGCGGATGCTTCTCT |-----|
137612
137637 ATTTTATAGGCTTTTTCGTACGTGACGTTAATCAGTTTGAATATTGGACAGGGTG |-----|
137696
137697 TCGTCTTGGCGCACGCAAGTTGGGTAGCGGATAGGTGCTGTAAATCGGGGATGC
137701
...
138063 GTGGGAATTTAGGGGATTTTGGGGGATTTTGCAAAGGTCACAGCGGCAAAATCGCCACC |-----|
138113

138123 TTCCCTCAACCTTCCGCTGTCCACACAGCAGAGCGGAAAAA
...
138271 TGGAGTAGCCGATGGCGTAAACGGCTTGGCCGATTAAGTCCGCCGCCGATCGCATCA |-----|
138321
138331 TACTCTGCGGGGAGATATAGCCTGAACACAGACTGGCGGTTGCATTCGCTTGGGGATGC |-----|
138375 138383
138391 CCGCACGGCGCTTGGCCGTCCACATTGTGATGGGTTGTCAAGTTTGGCCGACGCTT
138451 TAAATTGTGCGACGTTTGTGTAACATCCCTCTTTAACGCCCGGGGCGACATTGTAATGC |-----|
138461 138485 138494
138511 GGACGGTTTGTGTGCAATCTTGTGCGGACATTCAACG
...
138552 CCTGTGATCCAGCCTTAAAGTGGGGCCATCTGCCCGCGCGGCGAAGCAACGCTGGAA |-----|
138602
138612 AACTGCGGCGCATCGAAGTGGCGCATGTGCAATTTTACCCTTGGGCTAATCTCGATG |-----|
138618 138632
138672 TTGTTGCCGGCATTTAA
...
138756 CCTGCACCCGCTTGGAAATCCAAATTCACACCGCTGTTTGCACCTGATTAGCCTTAGG |-----|
138806
138816 |-----|
138818 GCATCGCGGACATTTACATCATACATAACAGTGAATGTGCTTATCATCTTACTTA
...
139129 CGATTTTAACCTTGGTCTCTTGGCGGTGTCTTGTCTTGTCTTCCACATTAACAGTCTGTTT |-----|
139179
139189 TCGTATCTGCGCTCAAGAACTCGACTGTGTGTAAGTGGGACGAA
...
139380 GCAAGCTATCGGTCAAAAGTCGAACGATACCGTTACGATGAACCGTGTGCGCTTC |-----|
139430
139440 GTCCACAGCGTTCTTTCGGGAATTCAAGCCTTTGCTGTATGATGATGATGATTC |-----|
139478
139500 TTGCCGTTTGGCCAAACGATAAATTTTTCAGTTTCAACATTGATCAAGGCTGTGAGGCT |-----|
139515
139560 TTTTTCAGCA
...
139578 AACCTACTGCAATGTGTTTCAATGAGTGTGTTTGTGATTTTCAAGTTGTGCGGCT |-----|
139628
```

133047 TTGCTCGCCTTAGCTCAAGAGAACGATTTCTTAAGTGCTGAGACCAAGTGATCG
133048 133070
133107 GTTCGCTACTACTGCTACTGCTGCGGCTTCGCTTGTGCTGATTTTGTAAATCCA
133108 133138
133167 CTATATGCAATCAATATTTTGTGAGA
...
133213 TAACCTTAACCGAGCCACAGACACACCGGCGGTGCTCAACGGAAGCGGCGTG
133263
133273 AGCGGACTTGGCGCAGGATATGCCAATATCGTGGCTGCTATTTT
...
133365 GGCATATGCGCGCTGGCGGAGCGGTCAAGCTGATTCGGCGCTCGATGTGGGATTTGAG
133413
133425 TTTCACACCAAGCGCTGCCGTCAAGCAAAAGCGGCGATGCAAGCTTGAAGTACGTT
133434 133454
133485 AAAACAAACGCTGGTGTCAAGGTGTTGAGCAAAAGCCGTTGCCAAGAAACCGGCTGGAAA
133492
133545 CTG
...
133597 TGGCTTATGCACAAGCGCGGCGCATTTGTGAGCCCTTTTTCATCAGCAAGACACTG
133647
133657 ATTTGGCCTTTGTTTAAGACGACTAAATVGGGCAATCTGCCGCGGATTTGCGGACGGCATTTG
133673
133717 CGATGGAATTGG
...
133888 AAGAAACTCAGGCGTGGCAGACCGCGAAACGGCGGCGGCGCAAAACGCAAGAAATGCTGTA
133938 133941
133948 ATCATGAGACAGGTTGCGTTTACGAAAAGATGATTGGCTGTGTGCGCAAAAGCCGT
134008 GAGGCAAGCGAACAGGACACTTGGGCGGTTTGAATTTGGGAGGCGCAACTGGCCAT
134034
134068 TATCAGGAATGCTGAACGGC
...
134097 AACCAAAAGTGTGAATAGCAATGCGTATTTGGATATTTTAAAAAACCCAGGAGAGGC
134147
134157 AATGTGTGCACTCGAAACTGTGGGCAAAAGCTTGGCTGCGCGGCGGAGCGGTTAAGTTT
134184
134217 GTGATGCTGCCCGACCCGTCGGC

...
134251 CGGTATTTTGGGCAATTGTCGGCGGCTATGCGGTGGCGGCTTCCTTGTGACAGGTAAC
134301 134308
134311 GTCAGAGGTCGAGATGAATGATCTCTGAAACTGCTGGCGAATAACTGGCAACC
134313
...
135035 AATGCACTCGGTCGCGGAGTTTGTGCTGTGTAGCTTCGCGCTGAGGATGAT
135085
135095 TCATGGTGGCTGCTTTCAAAGCGCCAGCTATTCGCTTTTGGCTTTATATGATTTCA
135114
135155 AATACATAAGTTT
...
135362 GGCAAAATATTTAATGGTATCAAAAGATGACCTACCGTGAATTTAGTTGAAGCTGAT
135412
135422 GGCTGTGGCCATGCCGATTTGGAAATTTGGGCTTAAAGCCGCGACGC
...
135504 TTGATTAAGCAGGCAATTGATGACGGGTAGCGGATGATAGGATTTTCAGACGAGTTT
135554 135557
135564 CACCTTGAATATCAATTAAGAACTATGACACCTTTAAACGTGCGGTTT
...
135626 GCGTATTACTGTGTTTGTAAATGATGATGAGACTGAGATTTGCCAGCAATCGCCCTGA
135676
135686 CGGTACGCCGTCCGTATGCTATTCGCGGATGTGCCGTTTAAAGGGTTTTAAATGAC
135711
135746 TTTGAATTTGTTTCAAGAC
...
135798 TTTGGTCAATGCAATTTACGCCGCTGTAAGAGGCGGACAAGCTATTTGACGACCTTA
135848
135858 AAGAGAGCTGCACGCGGTCAAGCTCTTATCACACCAAGCAAGA
...
135956 AAGTTGAAAAAGTAACGAAAAAACTGACAGGAAAGCGGACAAATCATAGAGCGCCGA
136006
136016 TTTTGAAGCCTTGGCGGCTATTGAAACACAGACTGATCAAAAGCTGAAATAATCAGAAAG
136056
136076 AATGCAAGCGGAAATTTGGCAAAATCCGCCAAGACA

```
129951 |-----| |-----| |-----|
AACGCCAAGTACGACGAGCAAGATGAGACGACCTGATGTTGTTGACGTCGCAAAAT
129951 129977 129998
130011 |-----| |-----|
ACGTAAATCCAAAGCCATGCTCTTCGGCAAAAGCGGCCAAACCTACGCGATCGCGCAG
130022 130028
130071 |-----| |-----|
GCCAATATGAGCCCGGTGGACAGACACCCCGATCGCCGCGCAAAAGCGCAAGATGCCAATC
130131 TCGACCTCAAGCGCGCATGTGCCGC 130100
...
130775 |-----| |-----|
ACTACCCAATACCAATTACGCGAGGTATAGTAATTAATCTAAACCAAGTACAGCGTTGCT
130825
130835 |-----| |-----|
CGCCTTAGCTCAAAAGAGACGATCTCTAAGGTGCTGAAGCACCAAGTAATCGGTTCCG
130852
130895 TACTATCCGTAAT
130908 |-----| |-----|
GTCTCGGCGTGGCCGCTTTCCTGATTTGTTGTTAATTCATATATCGACATCGCCAAAC
130958
130968 |-----| |-----|
GAACCTTGTGATCGCGCTTTCGTTCTTGTCTAAACCAAAACCAAAACCAACACCCCA
130979
131028 |-----| |-----|
AAGTATGCGCCATACCTGTTGATACCTTAATAAACACAAAGGTGCGCTGCTGACGCG
131033 131069
131088 |-----| |-----|
AAAGCAGCGGGGCTCTGAAATCCCGCGCCCAAAAGCCATCGCGCAGAGCATAGCCG
131148 |-----| |-----|
TGATTTATGCGCAACCCGCGTCAGAGCATGATTTGCCAAATCGAGCGCGCTGACCAAAA
131153
131208 |-----| |-----|
CCGAGCGCCAAAGATGCCAAAAATGCCCGCTTCTTCGCACAGATGAAGGCGACGAAGA
131210
...
131358 |-----| |-----|
GGACTGCCGAGAAAAACCGTCTGCATCAGTTTCAGAAACGAAAGTCGAAAGCGTCAAAC
131408
131418 |-----| |-----|
AACTGATTTGCCCATTTTGACCGGCTGATTCAGATTTGACAAACAAATGACGACCA
131468
131478 |-----| |-----|
CCCAACGCGATTTTGACGGCAAAAGCCCAAGTGGCGGAGCAAAATCAAAAGCATCGGTTGCA
131538 |-----| |-----|
TAACGAGCGGTACGCTGATGGCGATGTTGCCCGAATTGGGGCGCTGTCCGACAAACGGA
131539 131557
131598 TAGGAGTCTAGTCG
...
131666 |-----| |-----|
TTTGGCGGAAGGCTGCGGTGCGTAAGGCACTGTATATGCTACCGTGGGACGACGCT
131716 131716
```

```
131726 |-----| |-----|
TTTGAACCGCTTATTCGGGATTTCTACCAACGCTCCCGTCCGAGGTAAGCGTATAAG
131754 131754
131786 |-----| |-----|
GTTGCCGTTACGGCATGTATGCGCAACTGCTGACGATATCGAATGCCGATGCGTGAT
131818
131846 |-----| |-----|
TATTTTGGCAAAACGATACCGCGCAAAAGGATATCCGAGCGCTTGATTGAGTTTGG
131858
131906 TATTTTGG
...
132007 |-----| |-----|
ATATGCGTTAATTGGTCAATATATTTGTTTATGACTTTGAAATGCGTGGCGTTGAT
132057
132067 |-----| |-----|
GCGAAATTAAGATTGCTCAAAAAGAGCTGGGCTTGATGACGTA
...
132140 |-----| |-----|
CGGATTAAGCGGTCTGTGGCGATATGATGTTTTCGAACTTGAGTCTGTTGCGCTGATA
132190
132200 |-----| |-----|
TGCGGTCCGACGGAATTTAAGCCTAAGCAAAAGTAAACCCACACCG
...
132335 |-----| |-----|
CACACGAATGCGCGCGGATGTTTGGTAAGGATAGGTGCAATATTAGAGCATACGC
132385
132395 |-----| |-----|
AGCTACATTAACGTGTTGCTGCTGCTTGCAGATTGCGGAAAAACAGAA
...
132485 |-----| |-----|
CATTTATTCGGGATACCGTGTGGACATTTGTGATGTCATCGACCTGGCAGCGGAA
132535
132545 |-----| |-----|
CAGCTGTCAAGCGGATTTGGCGGGCGCGGTTTAAATTTGGTAAGGCGCAAGGTGACAC
132556
132605 |-----| |-----|
GAGCGTTTGGCAATTTTGTGCGAAGCCATCGGCGAAGTGAACACATGAGCTGTGCAG
132606
132665 |-----| |-----|
GTATATGTTGCGGAGGAATTTGATGTCACGAGTGGCGCAAGGCGTTAATACAGTTGAGA
132706
132725 |-----| |-----|
AACCATAGTTTATCAAGAGTTTGTCAAAATTTGGCGCATATTGATTAAGAAAGCGGCTT
132785 |-----| |-----|
ATGGCGATGACGAAGCTATGCGCTAAATACGCGCATCTCTTCAGACAGGGAATATACGA
132787
...
132927 |-----| |-----|
TTCAGACGGCTTTTGTGTTTGGTTTGCAGGGGTGAACATCTACCGTTGCGGACGATGGT
132977
132987 |-----| |-----|
TAAAGACGGTTAATGGGTTTCAAAATGTTATAGTGATTAACAAAACCAAGACGCGC
133043 133043
```



```
124710 GGGAAACGAAGCCGTTGCCCTGACCGATTTCAGACAGCGCCGCCGACGAGATGATGCCG      |-----|
124756      |-----|
124770 AATGCGCTGTGTCACCGACATCAGCGGTGTGTGCAGCGAGTGTGCTGACGTTCCAAACGA      |-----|
124774      |-----|
124830      |-----|
124830 CATAGTAGCCGATGACGAGCGGAGAGACGATTAAGTGAAGTTCAGGAATGCTGCGG      |-----|
124861      |-----|
124890 GTGCCAGCGCGCCAGCCACAGTACCAACACGCGCGGATGACGCGCGCGGAGTTT      |-----|
124924      |-----|
124950 TCCACAGGGGAAAGGGTTTCGCTTCGGGCTTGGCGGAGCGCGGCTTTTTCAGACGGCG      |-----|
125006      |-----|
125010 TTTGCTGGGGTTGGCGGAAACTGAAATCGCGGAGCGGAAAGTGATTTTC      |-----|
125011      |-----|
125241 TGCCGATTACGAAACAATTCGCCCTGTGTGAGTTTCAGATTGCGCCGCCGCGCGCG      |-----|
125291      |-----|
125301 CCAATCGACGATGACGAGACCGGGTTTCATGCTTTCAGACATTTTGTGTAATCAGCT      |-----|
125308      |-----|
125361 TGG      |-----|
125396 ATGATGATGTCACATCTTTCGCTGCTCGCAAAAAGCTTCTCGGCGCGGCGATTAAT      |-----|
125446      |-----|
125456 TCGTGCCTCATCATCTTTGGCGTAGCCGTGCGCGCTGCGCGCGGATTTTGGCGGAAGTCG      |-----|
125458      |-----|
125516 AGTTTCAGGAACCTGCGGCCA      |-----|
125565 AGCGGATATCGAACGCGGCTACACTGCGCGAGCGAGTTTGCCTACCGGATCGCGCA      |-----|
125615      |-----|
125625 AACCTGCCACACCTGCACCAATCACCMAAACCTGCGCGCGGCGACCTTTGCTGCGGCGAG      |-----|
125685 TAAATTTGGCCGGTGAAGAAACGCGGAGGCGTTGGCGCTTCAATCAGCGCGCGTAGC      |-----|
125713      |-----|
125745 CGCTGATGTTTGGCATTCGAAGACMAAGCGTCCMAAGCCTGCGCGCGGAAATGCGGGGCA      |-----|
125770      |-----|
125805 CCATGTCATTTGCCAGCGGT      |-----|
125837 GCGGCGAAGGCTTGCACCAAGCTCGTTTGGCGGCGCACAGGAAGCTGACGATGCTC      |-----|
125887      |-----|

125897 TGTCTTTTGTAGACAGCGCAAGCTGCTCTGAGACGGCGCTTAA      |-----|
125966 CAATCCGCTGCTTTGTGCGCAACGGTTGCGCCTGCTGCTTGTGTAACGGGCAATGTCMAA      |-----|
126026 CTTGCCGCCAATCTGCGCGGCTTTGCACACGGTTTCAAGCCAGTTTGCCCAAGCAG      |-----|
126085      |-----|
126086 GCGAGGTGCGCGCGGTACAGCGAGCGCGGTTTGCCTGATATGACTCAGCTGGGATA      |-----|
126106      |-----|
126146 CCGATTTTCACTCTG      |-----|
126174 GGTATTTATATGATCTGCGGTTAAATTTAAATCGGGGAGCGGAGACAGCCGATC      |-----|
126234 CGGTTAAACCGAATTCACCTCAAAATGTTAATATTTTATGATAAT      |-----|
126504 CTTCCCGCTGCCCGAACCCTGATTTGCGCCGATGAAGAAAGCGTATGCGTTTATGTTAG      |-----|
126554      |-----|
126564 AAGGTTTTCTTAAAGCGGGGCGGAGCGGCGCTTGGCAGCTGTCAATTTGACACAG      |-----|
126592      |-----|
126624 CCGTGTCCCGGATATGATTTGCGGCAATCGGTTGTGATGTGTCAGCGATATGATTCGA      |-----|
126668      |-----|
126684 CACTGATTACCATGCAATGCTGATGAAATTTGCGGAAGCGTGGGTTTAAAAACAAG      |-----|
126744 TAGCGAAATTAATCCGAGGCTTCGATGCGCGGCACTGATTTGCAACAGTCTTACGA      |-----|
126759      |-----|
126804 GCCGCTGCGGCTGTTGGCGGATTTGACGAAAGATTTTGGAGCAATGCMAAAGGCA      |-----|
126865      |-----|
126864 TTTGAAGCTCTCGCCGCGGTGCGAATTTTGTGGAGCAATGCMAAAGGCA      |-----|
126925      |-----|
126925 ATTCTGCTGATGTCGCGGCTTACAGTTTTTTACGAAAGGCTGCAACAGCTCTGG      |-----|
126975      |-----|
126985 CTTGAATACCAACACGCCAATATTTTGAATAATGAAACGGCAGG      |-----|
127082 TTGTTGGCGAATACCGAGCCTCTCGGATTTGACAGCCGATCAGGTGTTGCGGATGGGC      |-----|
127132      |-----|
127142 GACGTCGGAAGATATTCGATGTCTCAAAAGAGGCGCATAGGCTGCTTACCGTGC      |-----|
127144      |-----|
127151      |-----|
```

```
120327 GTAGTGGCGCTTTTCGCGCGCTTTTGTCTTTTCATCTTAAGAGAGATGATGAATAA |-----|
120333 |-----|
120387 GACGGCATGTACTACCAAGTCGGAATAAATAG |-----|
... |-----|
120642 GTTCCCTGCTGAAAAAGTTTGAATGATGATTTTATTATTATCTGCGCTTCCG |-----|
120692 |-----|
120702 TACCGTATTAAATCGGGCGGCTCTGTTAAGAAATGAGCGCATGA |-----|
... |-----|
121233 TAGAAGCGGCAAAATTTGAGACATCCCTACGGCGCAAAAGCCCGAATCAGCCAAAA |-----|
121283 |-----|
121293 TAACCGCAGCGTTTCCCGCGCGCGGATTTGGGGGGTTCGCAAACTTCCCGCTTAG |-----|
121301 |-----|
121353 GCGC |-----|
... |-----|
121822 GCAGAAACCTTTACCGCGTCATTCCACGAAAGTGGGAATCCGGGAACCTTAACGTTACAG |-----|
121872 |-----|
121882 TGATTATCGGAACGGCTGAACCGAATGAATTGATTCGCGCT |-----|
... |-----|
122048 CCCGATTAAGTTTCCGACACGACAGTCTGATTCGCCCTTCGCGGGAATGACGATCCAG |-----|
122098 |-----|
122108 GTATTCTGACAATTCGCGGCTATGATGTTGAGCAGAAATCGAAA |-----|
... |-----|
122223 AAAAATACATAGACGAGATTAACCGCTGACACGCGAAACTTACCAAAAGCCAGCA |-----|
122273 |-----|
122283 GACACAGCGTCAATCCGACTGAGGGCTACGGCACAGCGGCTTGTTCAAAGGCAAAAAAC |-----|
122288 |-----|
122343 ATATGCTTTGACTGACTTGAATGCGCGATTGAGCGGTTTAAACCGGAAGGCGATTTCCT |-----|
122377 |-----|
122393 |-----|
122403 TTGAGGCAAAAGCGTTGATGTTTGTATATGCACTTCCAAAGCCAGAGATTTTGG |-----|
122414 |-----|
122463 GTATGACCGCGCTGCGGACATTCATTATGTAACGATGTGTTAAAAATCCGCAAGTGGAAA |-----|
122492 |-----|
122523 AATACTTGGCAGATTATCAGGCACA |-----|
... |-----|
122846 TTGCTTGAAGACAGCGGTTTGTACACCTGCGTAGCGGTAATCAACGCGTTTGAA |-----|
122896 |-----|
```

```
122906 |-----|
GACGACGACTTCTTTTGCTTCCATACTTCCAACAGCGGCATACCCGCAATCGGGCTGTCT |-----|
122906 |-----|
122909 |-----|
122966 CGGGTCGGTTTGGCGCGCAGATTGACGGTATCGTTGCGCAACCAATCAACAAAGCACATC |-----|
122996 |-----|
123026 GGTTCGGGGAAGTCGTCTGTTGANTTCTCATTTCCTCAAAAGCATGTCGAGGACTTT |-----|
123038 |-----|
123041 |-----|
123086 GCTTCGGCGACAGTACGTTCAATATACCGGGCAGCGCGGACAGAGGAGTATGCC |-----|
123128 |-----|
123146 GAACGTAATCTTCCGTGCGGTTTAAAGCTCGG |-----|
... |-----|
123396 CACATTAATGTAAGACAGAAATTCGCGCTTGAGCCGACACAGCGCGGTACGATGAGC |-----|
123446 |-----|
123456 AGTTCGTTGGAGAGCATTAAGCTTGCCTGCGCTGCGCGCCCGCCAGCGGAGTAGAGTTGA |-----|
123458 |-----|
... |-----|
123601 GCAGATGAAGCCGCTGCCGTTAATGCCGACGAATACGACGACCAACAAAGCATACGG |-----|
123651 |-----|
123661 CAAGTGCAGTGCCTGAGCTTGTGTTGGCGGCGCATGTGCAGCGG |-----|
... |-----|
123777 AAAATACCAAAATAGACTGACCCAGATGATGATGTGATATGTCGAACGTTGCC |-----|
123827 |-----|
123837 GACTCGATGTAGCTGTTGAAGCCGACCAAAACCGCTGCCAAGCCGA |-----|
... |-----|
123924 ATTTCTACTTTTTTGTGCTTGTAGATGCCGATTGCGGCTCCGATGAGCATGCGATGATG |-----|
123974 |-----|
123984 ATCCAGCCAGATCCGCTGGGTATTGTGCAAAAAAACACTTACAAAAAGGCGCCATA |-----|
124031 |-----|
124044 CCGCGATFACCGGAATAGACAGCCCTGTTGGCGGTTCTGTTTGAACAGCCCGCAGT |-----|
124046 |-----|
124104 GAGAGATGAATAAATTTGGGCAAGATATACGCCCTGTTACGAGTCTTGAAGACATG |-----|
124127 |-----|
124164 GAAATCTCCGATTTTGA |-----|
... |-----|
124650 TTACAAAGAAAGCCGCGAAGATGTGATGCTGGCAATCAGAGTGCMAAACAGCAGCA |-----|
124700 |-----|
```



```
116938 CGGCGGGGATGGCGGCAAAACCTTCATATATTTGACGACTTCGATGTTGTTGCCGAC |-----|
116998 CGGACGGCGGAGAGTTGCCAAACAGGAG 116972
...
117130 CGGCAATCGGGTTTATGTTGCGGCCCAATCACCACAGTGCATGCGCATGTCGCCCTTG |-----|
117190 TACGCCGTTGCGTGGTTCCGAAGATTTTCAACATATTTGGGTC 117180
...
117556 CGCGGATATCGTATTTGCTCAACACGCGCGCCCTGTATGCGGAATGTGCGCTT |-----|
117616 CGCTTCAAGCGGCAATCTGCCGACGGGAAGCATGATTCAG 117606
...
118023 TGTATCAATGCCATCAGGCTTACTTCTTCGANTATGGCAACATGTCCTGAT |-----|
118083 GCGCATTTAGAACACGACGATTAACCTTATATTTGCGTCCCTAAGACAGGAGAT |-----|
118143 AACAAAAATTACGTCCTTACTTCTTACAGTAACAGGCTTTTTCGCCGTTTGTG 118136
118203 AGGATT 118153
...
118365 TCTTCGAATGTTACGAACGTACATACGAGCGCAAGAAACCTGTAGGCGTTCCTC |-----|
118425 TTCCGCGGGGTATACATTCGCTTTCATTTGATCAATTAAGCTTTTCATTTCATGAAA 118415
118485 AAACCTTTTTCGATTAATACGGCGTTCGTAGTCTTTTGGAGACGAATTTTATTC |-----|
118545 GCGCGCGGTCCATGCT 118505
...
118564 CGAAGAATTTTTCGTTTCGATATCTACAAAGTCCAAGSTTCGGCGCGCTTTTAA |-----|
118624 TGAGCGCTGTTGTTGATGAGTTCCGAGACGTCCTATACGGTCGTCATTTTGGGGG 118614
118684 ACAAAACAATACATTTCTTTTCGAACGTACCGGACCGGTTGCGGCAAAAGAAAG |-----|
118744 CTGGAATCAGACTTTTCGATTCCTGACTAATGCAAT 118727
...
118884 GTCATCAGTCAAACCAAGCGAATGTTGGGTCAGATTGGAGAGAACGATGAAG |-----|
118934 118934
```

```
118944 CCAAAATGACCAAGGCAAGACCTATGATTCGGTCCGGTAAATGCTCAAAATATGTCG |-----|
119004 GCGTCTATGAAAAAGCAAGCAGTTGGCGATTAACAAACACATGAGCGGATTTGAAA |-----|
119064 TTGAATTCAAAGCAAAAGATATCGTTATCCCTTTCCAGATTTTGCAATTCGGGGCAAT |-----|
119124 ATTTGCGCGCGCATATCCGATTTGGCAAGATTCGTCCAAAAGCAACGCGCATACAG |-----|
119184 CGGTTAAGGAGATTAAGTCAAT 119151
...
119523 ACAAAATTCGCAAGACTACCGATTTCATGGGCTTAAAGTGTCGCCGAAGACGTTAAT |-----|
119583 CACACAAGGAAACCAAAATGACATCCACTTCAGGCGCACATCTGCGGGCTTAAGAA |-----|
119643 ATTCAAGGACAAATCGAAGGACAGGAGTTTCAGTATTTGCCGCTGATTTGCGCACACC |-----|
119703 CTTAGACAGCTCCCAAGGCAACGCATTTGGGCACTCTACTGAATACGATTTGCGGG |-----|
119763 CTCTGCCAATTTTCGAGCAGTTCCGAACGCCCAATTTCCCATCGAAGCAAACTGAACT |-----|
119823 AGAAATCGTCACTACGGGCAAAACCCAAAACACTGAAAGTCATGCTTTTCAACCCGTTAA |-----|
119913 TTGTCCAGTCGATCAACAGGGGACTTCTGTATCTCTCTCCTGAACGCGGACATCG |-----|
119973 GACATACCAAAATTAATCAACAATGCCGATTAATTTCTAGACTTCGAGAGAGGATTAACG |-----|
120033 CAGGTTTGAAGAAATCGGCAACCAATAGAAATTTGTGATATTCGATTTTGAAGACT 120022
120093 GATTTTGGATGTTGCGGGTCTGTGAAGAAAGCTCCATTCATTTACGCCCAACACTTT |-----|
120153 TTGAAGAAAAATATCATGAAATTTAT 120112 120123
...
120207 CTGGCTGTTGTAACGCTGCTCCCTGCTTTGGCGCAATGCAAAATGCAAGTTGCCC |-----|
120267 GATACAGCAAAAGCGTTTGGAAAGCCGCAAAAGCGGATATGGAAGCCGTTGATTT |-----|
120279 120279
```

```
113561 TAGTAGTCCAAATGGCATCGATGCGGTGCTCAGCGAGGCTTTCAGTGTGAGCGCCGAGG |-----|
113621 CCGCCAGCCGGAACCGGAGGCGCTTTTGGTTCGGCTTGGGCATGATGAGGCGGTGGCCG |-----|
113681 TAGGGCTCTTCCCTGCGCGGTAAGGAAGGCAAGTTTTCAGAGGCGCTGCTACATATGTCG |-----|
113741 - |-----|
113742 CCGGAACGCCGATGCGGTAGCGCGGTGTCGGTTTGAAAGTAGATTTCGCCCTCCGCGCCG 113751
113801 TAGAAGTCCGCAACGGATTTGTTTAGCGCACGAGCTTCATTTGCTGCTCTTCATATG |-----|
113861 GATTTGGGGCCGCGTCCGTCGTTTAAAGTTTGGGCGTAATGATGTTGCCGAAGCGGTTG |-----|
113921 CCGTAGAGTGCAGATTGTATTGCCCAGCGCTGCTGCTAGCCAGCGAGATTTCGATA 113950
113981 TTGTTGGAACGCTCTTTGTTGTAAGTGTGTTGCCGACTTCAAAAGGTGTGTTGCGCAGG |-----|
114041 TGTTCGCCGTGTCGTAACAGTCTTTCGCTTGACGGCAGGCGTTCCTGATGGAGGCGGTC |-----|
114101 AGGCTGAGTTGTGTGTTGTGGCGTAATACAG 114078
...
114416 CCCAGCTGCCCTTTCAGAGCGGCTTATGGGTGTGGCGCACTGATACGATACGGCGCTTTTGC |-----|
114476 GTTGTGTTGTAATAAAGTTTCTACTGATGCGCTTTTTCGTCGTGGGCTATGTCG 114505
114536 TTGCGTTTCAGGTGTACGCGCAGGGCTTCAAAAACCGGGGAACGTTGCTTCCATTGCGCG |-----|
114596 CGGAGTTGCTAGCGCTTTGTTGCGCAGGTCTATCCAAGTTTGCCTGTGGGCAATGGCGG |-----|
114656 TGTGCATCATGCTGCTCGTGAAGCGCGAGCTCAAGCCCGGATTTGCTGTAATCATGTCCT |-----|
114716 TCTTCGTCACACAGGTGGCGATTAAGCTGCAAAATGAGCGTTTGTATCAAACTCTTTTGC |-----|
114776 CAGATGATGTCGCGCTGGCAATCATGTAATTCGTGCTGTGGCAGAGCAACATATTGG 114750
114836 TCGCGACGCTGCTGATGTGCGCGCATAAAGCTTTTTCGCCAACCCAAAGACAGCCCG |-----|
114844 |-----|

114896 ATGCTGCCGCTTTGGCAATCGGCGTGTGTCGGCAGGCGTTTCAGATTGCGGTAAACGC |-----|
114956 GGTACGGCGTAATCCCGCGATTTCGGGTACAGC 114933
...
115166 CTGTACAAGAGGCTAACCGGACCGCGCAGAGATTTCGACCTGTTGGCACAAGGCGGCTGTC |-----|
115226 ACCATGATTTGATGCTGTGAGAGAAGTCCGCCATGTCGCCCGCTTTCGCCGTGATGTTTC 115261
115286 AACACTTTAATCCGTCGCTGTTGACCGC |-----|
...
115367 GGTACGCGCGTCTAAAGCATCACCAGTTGACGGCTTTTGTGCAAGGTGTCGCGCGCTG |-----|
115427 ATGATTTTGTGAGAGCGGTAGAAAGTGTGCAGACAGCCCGAAGTGCGCGCGGACGCTT 115417
115487 TTGCGGACGAGCGCTGACCGCTTTCGAAGCCCACTGATGTGCTCAGTTCCATGCGCTTGGAG |-----|
115492 115492
115495 115495
115547 AGGA
...
116025 ATTACAAATGCTTTTGTATTATGATTAATTATGACAGTTTTCATTAACAACATGCTCCAGT |-----|
116085 ATTTTACAGAAATTGGGGCAGAGATTGCCGTGCGCCGCTTACAGATGATTTTACGTTGAGAG |-----|
116145 AAATCGAGCAATGTAATCCGCAATATCTGTTATGCGCCCGGTCCTGTTGCGCCCAAG |-----|
116205 AAGCGGGATTTCCGTGGCGGCGATGCCCATTTTTCGCGCAGGCTGCCGATTAATGGCG |-----|
116265 TGTGCTCGGGCATCAGACGATAGCGAGGGGCTTTCGGCGGCAAGATACGCGCCCAAAA |-----|
116325 CGCTGATGCACGGCAAGGTGTGCGCCGTGTACCATTCGGGCAAGGATATGTTTAAAGGTT 116328
116385 TGCCCAATCCGGTTA 116344
...
11663 CGAATATTGAACGAGGAGCAAAAAAATGATTACACCGCAACAGCCATCGAAGCATTAAT |-----|
116723 CAGCAATAAGAGTTGTTTACGATGAATGAAATACGCGATTTGATGCGT 116713
...
116878 GTGAGTTTGCACAAAGTCCGCTGAGAAATGCAGAGGGGCTGTGTCGATATCGTCGGTA |-----|
116928 116928
```

```
110216
110226 CGCGCTCGTGGCCGTGCGGTACGCGGTTTCAAACTTTCCGCCAG
...
110412 GCAGTCCGGCAAGAAATCGAACAACTGCCCTTACGCTCTGTAACCGCGCGCGCAAGAA
110462
110472 TCATCAGCGCGGTGCTCTCCGCGAGTTTAAAGCGGCATCAACCAT
...
110628 CGGTGGAACGCGGATTTGTGGCTGGTGGATTTCTGCTCGATGAAATCCGCGCGCAGAGC
110678
110688 CTTGCATCAAAATCGACATCGGGTTTGTCCATCATCTGCAAAAACCTGCCGCGCATAGCGCG
...
110748 ACAGGCTCTCGACATTAACGGGCTTGCGCTTGGCATTAAGTATCAAAAGCGTAGCGAGC
110766
110795
110804
110808 ATTGCGCGCTGCCGACAACTCCTGTATCCACACGAGCTTGTGGCGCGGAATGTCCAAAT
110809
110868 CGACGTTTTCMAAATTATGCGTGGCGCGCGCGGATGCGGATGATGATGCTGCTGGCGCGTGGC
110869
110914
110928 AATGTTGGGATGATGTTGTCACATAATGATGCCCGCTGAA
...
111084 CCGCTGAACAAAGATGACAAATGAACCGTCTTTACCCCGCATTTATGCGCGGTG
111134
111144 AGGCGTGGCGGATTAATTGGCGCGGCTTGGCTTGGACCTGCTGCT
...
111365 GCTTTGAAAAATCAGTATTTTATGAACGTGTCAACGTGCAATTCGCAAAAATTCGCTGCC
111415
111425 GATTGTACGGTAAAGAAAGTGTCTATTAACAAGCAAAATTCGGAATCGGGATTTGGAC
...
111485 AAAGCCAGCAGGAAACGAACGTAAATGCGGTGTTGGCGACTACGCGCTCCGCTCGTGAA
111520
111545 ATTACTTTTGTCAAGTGGCGGTTTGGTGG
...
111577 GCGCCGATTTTGTGTACACCCAGCAGGTGCGAAACTGTCTGCGCGGCGCAAGCCTATGG
111627
111637 CTTTATCCGTTTGGCTCGCGCGTGGATATGTATCTGCGTGTGAGCGCAGCGCAAGT
...
111697 GGGATTTGGGATTAAGTAAACGCTGTCAAGCACTGATTTGGCGCGCTTGGCGGCTGACTGC
111697 111704
111757 GCC
...

111976 TTGCTAATGTTACACCGCTTATGTTTCAGAAAAATGGGAATCCGGGAACCTTAACGTTACAG
112036
112039
112123 ACGGAACCTTTAGCGCGTCAATTTCCACGAAAGTGGGAATCCGGGAACCTTAACGTTACAG
112173
112183 TGATTTATCGAAACGCGCTGAACCCGAAAGAAATGGAATTCGCCGCT
...
112285 CCTTCCGCTCATTTCCACGAAAGTGGGAATCTGAACCCAAATGCTAAGCGGATTTAT
112335
112345 CGGAACGCGTGAACCCGAAAGAGACCGGATTTCCCGCTGCTGGGACGACAACTCATTA
112390
112405 GTTACTTAATAAATTAAAAACAGAAACCTTTACGCGCTCATTTCCACGAAAGTGGGAATC
112465
112474
112525 CGCGTGGCGGGAAATGACGACTTATTAATGTTACCTTAATAAATAAAGCAAACTTTAC
112540
112585 GCGGTCAATTTCCACGAAAGTGGGAATCCGGGAACCTTAACGTTTACAGATTTATCGGAA
112621
112645 CGGCTGAACCGCAACGAATTTGATTTCCGCGCTGCGCGGGAATGACGACTTATTAATGTTAC
112687
112705 TAAACTTAATAAACAAGAAACCTTTACGCGCTCATTTCCACGAAAGTGGGAATCTAGAAC
112765
112776
112825 CGCGGAATGACGACCTCATTAATGTTACCTTAATAAATAAAGCAAACTTTACCGCGTC
112834
112885 ATTCCACGAAAGTGGGAATCCGGGAACCTTAACGTTTACAGTATTAATCGGAACGCGCTG
112915
112945 AAACCGAAGATTTGATTTCCCGCGCTGCGGGAGATGACGACTTATTAATGTTACCTTAAC
112981
113005 TTAATAAACAAGAAACCTTTACGCGCTCATTTCCACGAAAGTGGGAATCTGAACCCAAAT
...
113065 GCTAAGCGGATTTATGGAACGCGCTGAACCCGAAAGCAATTTGATTTCCCGCTGCGCGGG
113070
113125 A
...
```

106525 CTATTTCATTGTCATTCCAAAGTCATTTCATCAGCTTTT
...
106591 GGGGAACCACTGCATTTTACAATTCATTCACATCAATATCATTTGTTAATGCTTTCAT
106651 TTCTTCATAATTTTTCAGCTAATAAATAATTTGCTGAATTAATCG
...
107199 TTGCGTTGAGATTGGTTCAACTCTGGCAATAAGGGCTTATATAGGAGAGAGCATTTGTT
107259 TTGATAATTGCATTATATTTCATCAATCTTTTCATGGCAATAATGA
...
108149 ACCTTGGCCACATTCTCTCCGGGTACCTGAAGCAATAATCCTCCCTCCGCATGCGCCCCC
108209 TCGGGCCCTAATTCACACATTCATCTGCAGTTTAAATCAATCA
...
108272 ACAATGCAGTTGCTTTGCCCTTTCAGACGGCCCTATGACTTCCAGCAGCAGGGCGATGTGCG
108332 GCGAAGTGCAGCGCCGGTGGTGGTTCTGTGAGGATGTAGAGCGCTTCTGCGCGGTGTGCGGT
108392 TTGGAGAGTTCCAAAGCGAGTTTGACGCGCTGGGC
...
108473 CCGAGACCTACATCCATCAGAGTTTGCAGATTGCGCGAATACGCTGGGGACGGCGTGGAAAA
108533 AATTGCGAGGCTTCTTCGACAGGTCATGTCGAGAGCTTGCTGATGTTTTTTCCTTTGTAT
108593 TGTATTTCCAGGCTTTCCGGCTTGTAGCGTTTGACGTTTGCCTGACAGACTTTCGAGGGGACGTAC
108653 ACGTCGGGCGAGAAATGCAATTTGCAGCTTAATACAGCCGCTGCGCTTGGCAGGCTTCGCAG
108713 CCGCCGCTTTGACGTTGAAGAGAAATCTGCGCAGCTGTAGCCTGTTACACGCGAGAGG
108773 GGTACGCTGGCAAGAGTTCCGCGGATGG
...
108815 GTGTAGGTGGCGGAGATTGGAGCGTGGGGTGCGGCCCATGGGGGATTTGTTGACGCTTGATG
108875 ACTTTGTGAGGTGTTCGAGCGCCGCGATGTCGCTGATGGGGGCTTCTTGTGGCT
108935 CGGTTGAGTTCCGCGGCGGTGATTTTGGCGAGGCTTCGTTAATCAGGCTGATTTGCCG

110166 GGCAGGCGAAGCGCGGAGAAACCAATGTTCTTCGCCGCTGTCATCTTCATCCGACGCG |-----|
108995 CTGCCGGATAGCCCGGTATGACAGTAATCAAAACCGAGCGGACAGTTCCAGGGTAACTGTT |-----|
109048
109055 TTGAGGTGTGTGCGCGCGCGCTTTTGAGGACAGCATTCGCGCGGGAT
... |-----|
109133 ATGATTTTTTTGCGCGTGAAGTAATGTCGCCGTAAACGAGATTTTCGCATTTGGCGACGTTT |-----|
109183
109186
109193 TCGGGCGGTGCGGCAATCAGTACGTTGCCCGCGGTGTTCCGCTCGCGCGGGGCCCAATACG |-----|
109253
109257
109276
109313 CCCAATTCGCGCAGCGGTTTGAGGGGTGGCAAGCAGCGCGGCGGTGCTGTGTCGCGCGGCTGTGACG |-----|
109326
109351
109373 CCGATGAGAGGCTCGTCCACACATACATACGC
...
109457 CCGCCGGAAGAGGTTTCGGCGGAGCGGAGAGATTGAGTAATCCAGCCCGCATTAATC |-----|
109507
109517 AAAAAAGCGAGCGCGTCCGTTGATTTCTTTGAGATTTTTTCGGCGATTTGTTTGTGTTGTTG |-----|
109558
109577 CCGTCCAAATCCAGCGTTTCAAGAAATCGGTGGGTTTGTGTCAGCGGCCAGCGGAGACT |-----|
109589
109637 TCGTGCAA
...
109726 GTATTCGCGCAATTTTTCGCGCAGCGTTTCGTCGTGCTTCGCGTTTCGCGGTACGCGCGTTTCGAG |-----|
109776
109786 ATTGGGATGATGCTTCGAAAGCGGTGCGTTCGCGTTGAAAGTGCTG
...
109905 CAGGTAGGTTTCCCAAGCAGCCTGCACATCAATAAACCGTAATGCGGTGCGACAGCAGCTAA |-----|
109955
109965 TCATTTGGAATAGAACTGTTGCGTTTGTCCAGCGGTCAATCGCGGCCGGAAGCCACGCG |-----|
110021
110025 ACAATTCGGGGGTGGGCGACACTTTTTCGGGGGTGCAAGAAATTTGTTGCCCAAGCCGT |-----|
110026
110038
110085 CGCAAGTCG
...
110166 GGCAGGCGAAGCGCGGAGAAACCAATGTTCTTCGCCGCTGTCATCTTCATCCGACGCG |-----|

```
103731 TGCGGACAAATCTCAAGAAAAAGAACGTGAACTGAAGCTTCTGA 103721
...
103790 TATGTTGCCCTGGAAGGCAACATCGGCTGTATGTTAAAGGTGGCGGTTTGGCGATGGCC 103840
103850 ACCATGACATCAATCAAACTCAAAAGCGGTCACCTGCCAACTTCCTGTGACGTTGGCGGAC 103898
103910 GCGGCAACCAAAAGACCGGCTGTTGAAGCGTTCAAACTGATTTTGAAGACAAATCCGTT 103936
103970 AAAGCGTATTGATCAACATCT 103936
...
104004 TACGTTGGACATGATTGCGGAAGCCATCGTGCGACCGCTTAAAGAAATCAACGTCGACG 104054
- |-----|-----|-----|-----|
104064 TTCTGTGCTGTTGCTGTTTGAAGGTAAACAAGCGGCAACCTCGCGGCAAAATCCTGAAACG 104060
104124 AATCAGGCGCTGAACACTTCTGCAGACGGCTGAAATGACGCGAAGCAAAAAATTTGTTG
104184 CAGCGGTAAAGCGCTTAAGAGAAAGAAATGAGCGTATTGATTAAATAAGACCAAGATA 104192
104244 TTGG
...
104272 GTACTTTCACACTCGAACAAGCTCTGGCTTACGGCATAAAGTTGTGGCGGCGCTTACCC 104322
104332 CGGGCAAAAGCGGTCAAAACCCACTGATCTGCCGTATTTCACAC
...
104404 GCGCGATGATCTGTAAATTTACGTTCCGCTCGTTTGTGTGAAATTCCTTCCTGAAG 104454
104464 CCGTTGATTGAGGCTAGGCTGTGCTTGTGATTACCGAAGCGGTACCGACTCTGACAA 104487
104524 TGCTCAAGCCAAACGTTATTTTGAAACCAAGGCAAGGCAACGCTTTGGTGGGCCCTA 104536
104584 ACTGCCCC
...
104704 CACAACCAACCAACTGGGCTTGGGTCAATCAACTGTATCGGTATCGGCGACCCGA 104754
|-----|-----|-----|-----|
104764 TTCCTGTTATGAACCAATTCGACGCACTGAACCTTTTCCAGAGACCCAGATFACCGAGC 104820
- |
104824 CCATCATCATGATCGGTGAAATCGGCGGTACTGCGGAAGAGAACAGCCGATATACATCC 104825
104884 AATCCAAAGTAAGCAAAACCTGTTGTGCGCTACATCGCCGCTGTACCGCTCCTAAAGCA 104916
104944 AACGCATGGGCCACGCCGCTGGCATTTAT
...
104977 GCGGCAAAAGTACTCGCAGGAAATAATCGCCGCTTTGAAAAACCGGATATCGATACA 105027
105037 CCGGCAAGCCCTGCCGAGTTGGGCACTACCAATGCTGGAAGTGTGAA 105027
...
105360 AGCAGCGCAATTTTTTGTGATAGATTGGAATGATGTAGTCCGCGCAGCGCTGCAA 105410
105420 ATCCGTTTGTGCGCGACGCTGTGTCGTTAAAAACGCCGTTGCCAAATACGCGTTGCTG 105452
105480 CTGCGGCTGTTTGC CGCGGATTAAGTTT
...
105523 TATGGGGAATTTGTTTAAATCAAGCCCGCGAGTACGGAATTAATCCGATCGCGGC 105573
105583 GGCAGCGTCCAAATGGCGTGAATATGTTGCGACGACGATACGGGCGGCTTCAAG 105586
105643 GGATATTGTTTTCACATCCATATTAAGCCCGACGCGAAG
...
105974 TATGCAAGCTACGGCTTGCTGCTGCAGAACTCCTGAAGAAAAAGCCAAAGAGCGCTG 106024
106034 AAGAGCTAAGAGAAAGCAAAAGACGACGAGGAAGATGCCAAAGAAAGCTCTGAAAGG 106090
106094 CTAAGAGAAAGCTAATAATTCGATAGCTTTTACAGAAAAACGATTCCAA 106095
...
106321 GGTGCTTGTATATAGAGCTACGGCTTGCTTATTTTATCAAAAGGACGCGCTAAGT 106371
106381 GTTCAGCATTTGTGATGCCGATATAAATAATATATCTAACAGCA
...
106465 ACAAATGACTCGGCAATTGCTGTAAGAAATATCCGATATTTGAAGCTCTAAGCGCCATT 106515
|-----|-----|-----|-----|
```



```
97534
97568      |-----|          |-----|
      TGGCGTTCAACCCGTCACACTTGGAAATCGTTAAACCCGGTACTGCAAGGTTCTGCACGCG
      97569      97594
97628      |-----|          |-----|
      CCAAGCAGAAAGCGTTTGGGCGGAAAGGCGCGGATTAAGTCTTGGCCGATTGATTACG
      97637      97687
97688      |-----|          |-----|
      GCGACTCGCATTTATCGGTTTGGGTGCACCAAGCAGCATTCATCTGTCTAAAGCG
      97692      97722
97748      |-----|          |-----|
      GCGGTTATACCAAGCGCGGTACGGTTCATATGCTCATCAACAATCAATCAATCGGCTTACCA
      97777
97808      |-----|
      CTTCGATATCCGGGATACCGGTTG
      ...
97841      |-----|          |-----|
      ACTGATCCGATATGCAAAAAATGGTTCCGCTCGGTTATCCATGTGAACGGCGATGATC
      97891
97901      |-----|
      CTGAGCGTGTGTTGCTTGGCATCCAGCGCGCTTGGATTACCGCAAAAAATTCATTAAAG
      97961      |-----|          |-----|
      ACATCGTATGACGCGTTGTCTGTACCGTAATAGGGCCACACAGAGGGCGACGATCCGA
      97972      98008      98011
98021      |-----|
      CCTGACCCCAACCGATGATTAACAAAAAGTATCGCAACCCGCGG
      ...
98137      |-----|          |-----|
      ATCCAAGCCTACCGTGATGCTTGGATTAAGGCGACACGTCGACCAACAGCTTGAAC
      98187
98197      |-----|
      AACTTCACAGCAGCACAAATTTGACTGAGACAAATACCAAGSTAAAGATTGGCGGAAAC
      98257      |-----|
      ATCGAAACCGGTTTGCCCTGCCCGGATATTGAGCGTCTTACCGAATAATTTACCGCGCTA
      98288
98317      |-----|
      CCGGAAGGCTTTGCCCTGCATCCGACT
      ...
98489      |-----|          |-----|
      CGTTTCGCACCGCCACGCGTATTTGACAGATCAAAAAACGCGAATAATGGAGCAGCGTA
      98539
98549      |-----|
      CTTATGTTCTCTGCGCCATATGGGCGAAGAAATGGCGAGTTCCTGTTATGACTCCA
      98609      |-----|
      TTTTGAACGAAGACCGGTGATGCGGTTTGAGTACGGCTTTGCCCTGCTCCGCGCTGACA
      98631
98669      |-----|          |-----|
      AGCTGACCATTTGGGAAGCTCAATTTGGTACTTCCGCCAACGGCGCGCAAGTGACGATTG
      98721
98729      |-----|          |-----|
      ATCAATTCCTGTGCTAGGCGGAAACCAATAGGGTGCCTTGTGGCGTCTGACCAACATCC
      98739      98762
98789      |-----|          |-----|
      TGCCGACGCGTTACGACGGTCAAGGCCCTGAGACCTCTCTCTGACGCGGTAGAACGCGTGT
      98840

98849      |-----|          |-----|
      TCCAACTGTGTTCTGAGAACAAATATGCAAGTCATCATGCCATCTGAAAGCGTCAAAATGT
      98895
98909      |-----|
      TTCACCTCTTGGCAACGTCTTAGGTTTACCGCAAAACCGCTGTGAATTTTCATGT
      98921
98969      |-----|
      CCAAGCGCTGTTGGGATTTCAAGAGGCGCATGACCCCGCTGGAATACTTACCGAAGTT
      98972
99029      |-----|          |-----|
      CGACTTCCGTCGCGTTATCGGCAATACCGCAACCGCGCAGCAGCAGCGTGAAC
      99049      99073
99089      |-----|
      GCGTGTATGTGTGCGCGTCAAGTTTACTATGACTTGA
      ...
99266      |-----|          |-----|
      AAGAGCCGAAAAACCAAGCGCGTTCTACCAATCCGCCACCGCATCGAAGACGTTATCA
      99316
99326      |-----|          |-----|
      GCGAAGACCAAAACTGTCTTATGCTGCTGTCGCAAGCAGCGCATCGCCTGCAGTGGGT
      99353      99369
99386      |-----|          |-----|
      ACTCAAGCAAAACACTTGCTCAATTGAATAAATGTTGTAAGAGCGCTTGGCGTTGTAA
      99427      99435
99446      |-----|
      CCAAGCAGCATTCGCTGTGAGGCTGCTCAGCGGGAATGCCATAT
      ...
99839      |-----|          |-----|
      CCGCTCTGCGGAAGCTGCCCCGCGCGCTCCTGCTGTGTCACAAACCAACGCGCTA
      99889
99899      |-----|
      TGCTGCGCGCGCCCAACTGCGCTCAGAGAGCGGTGTGACGTGA
      ...
100056      |-----|          |-----|
      ACGTCTGAAGAAGCGGTACCGATGAGCGCGCTGCTGCTGTTGACAGACGCTTT
      100106
100116      |-----|
      GCGTTCACACAAAGAAACGCGCATTTGACTACATTCACGAGTCAACATGAACCAAT
      100132
100176      |-----|
      CATGACCTTGGC
      ...
100418      |-----|          |-----|
      ACATGAACAAGCAATTTGTTGATTACGCGAAAAAAGCCAAAGACGCAAAATCGCTATCG
      100468
100478      |-----|
      AAGATCTGACCGGCGGTACATTCAGTATTAACCAAGCGCGTACTTT
      ...
100705      |-----|          |-----|
      GCCATTAAAGACGCGTTGGAAGACCAAGCCCGCTGTTGTTGATCTGTAATGTTACG
      100755
100765      |-----|
      ACGGCGTTTATTTATGTAATGAAGCGCTGTAATTTTATAGTGATTAATTAATAA
      100825      |-----|          |-----|
      CCAGTACAGCGTTGCCCTGCGCTTGCCTACTAATTTGACTGTCTGGCGCTTGCCTGT
```

```
94364 ATATGACCCAGTTCTTCAAAACAATACCATTCCGTCAAAACCTTATGTTGTCACGATATAC |----|
94424 CGATTGATGGGACAAAGCGCTGTGCAAACTCAGCAGAGCGTTAAGAGTTGACGGTT |----|
94484 TGTACAGTGTATT 94442
...
94572 GCTGAATGCTTACCCTTTCATTGCGGACAGCCGTGATACCATCACTAATGACGTTTGA |----|
94632 TAAATCGAACGACCCATACCCTTGTTCCTGCTGCCACACCATTAATGAATGCGTAGAGT |----|
94692 ATG 94639
...
94730 AGATTAAAGAGATTATGTTGAAACGGGCGCTTAAAGAAATGATGTTTTCGACGATATG |----|
94790 CCAAGAGGAAATCCGTTTTCAAACCCGCGGGATTGTGAGTTAGATTATATCTTCG 94780
94850 GCAGGTTTATGMAAAAGAAATTCGAGCATTTTGAGCCGATTAAGAGCTGTCCGAGTTTCCG |----|
94910 AATCCTGATTTCAAGATCAAGAATF 94882
...
95261 CCGATTGTTTTCACCGCAAGCTGTGAGTCTAAATTAATTACTTACATGACGGGATACG |----|
95321 GCTTGCTTATTATTCGCGGATACCCTCATCGAGCAGCTGGCCGAAAA 95311
...
95473 GAACAGCTGACTTGTTCCTTCGCGGGGTTCCGCGGACGCGCATCCGATGGCGATGATG |----|
95533 GTCGCGGTGGTGGCGCACTGTGTGCTTCTACCAAGACAGTTGG 95523
...
95774 TGGCAGCGCGGCTCGACCGCANCTTATTTTCATGTCGACGACAGCAAGCAAAAGCCCTCAA |----|
95834 CTTCAACCGTCCGCTTGCGAGGGTCTTCGGGTGCGAACCCGTTGCCTGATATTGCTGCCG 95824
95894 GTATCGCTGCTGTGGGGTGTTCGACAGGCGGTGCGAAGAACCGTGTGAAAATGT |----|
95896
95954 TGGACGAAATCGCGGATGTGTCTAATGTTGCCGATACATGAAGGTGTGAAAACAGCA |----|
95965
96014 AATACCG
...
|----|

96091 TATGAGTTTGAAGAAATTTGGGCTTGGAAAGACAGTCGCAAAATTCAAATGGCGATGSA |----|
96151 TTGGAACAGATTGGCGTGAAGAACCCGTTCTTTATTCGAACGCAAACTGTATCCAAAGTCTC |----|
96211 GATTTCATTCGCGCATGCTGTCCGCGCTGGGCATCCGACCGAATGTTTACCCTC |----|
96271 ATCTTCGCCCTGTGCGCAGCGTGCGCTGGAATTTGCACTGCGACAGAGATGATTGCGAT |----|
96331 CCTTCGCTGAANAATCGGCCGCCGCCGCGCAGCTTTATACCGGTTGCGAAGCGCGGATTAT |----|
96391 GTGCCGCCAGGCGAGAGTAACATACATTATATTTGTCAA 96377
...
96833 CATCTGCCGTTGCGGCGGTGCGGATGAGGCATGCTGMAAAAGCAAGTCAGCGTTTTCAC |----|
96893 GCGTGAATTCGCCCTATCGTATCCAAAGCGGTGGTGACGCCCACT 96883
...
96967 GATATTGAAGCCCTCGATCCGAATTCACAGGCTGTGCAGATGCCGATATGCGCTTCAA |----|
97027 TTCAATATGGGCGAGGCGGATTTTGGCCCAATCGCGCAAACTGCCCTTGTGCCCAATCATC |----|
97087 AGCAACCTCAA 97042
...
97104 CTACTGGGCGACATCGCATTGGAATPATATCTATATTCCAATCCGAATCCGAAGCGCCGTG |----|
97164 GGTACGCATTTATTTGAAAGCGTATTGTCCACACCCGATTTACAATGCGGATCAAAAACG |----|
97224 CCGTATCTTGAAGAGATGACTGTGCGGAGACTTTTGAACGTTATCTGCATACCAATAA |----|
97284 TGTGCGTCAGAAACGTTTCGGTGTGGAAGGCGGGAAGCGCGATTTGCCGTTTGAACATA |----|
97344 CCGTATCAAAACGCGGTAAAGACGCGGTGGAAGAGATCATCATCGGTATGGCGCACCG |----|
97404 TGGCCG 97354
...
97448 GAGATTGTTGCGTGAAGGCGGTCGCCGAATCAAAATTCCTAGCGGAGCTGA |----|
97508 AATACCAATGGGCTTCAGCTCCGACATTCGCCACTCCACATGATGTCGATCAGCTTCTT |----|
```



```
91820 91810 91877
|-----|
GAGCGTGTATGTAGTAACGTAATTCACCGTGCCCATTCACGTTGGGATTTGGC
91821
91880 91882
--|
GATGCAAGCTGGACCGCGGTATTATGTGATTTATACCGTTGCACCTTTAG
...
91972 92022
AGGCATTTTTAGTCAAACTGGGTAAGATTTTACCCAGTGAAGCTTCATCGCGGTAT
92032 92022
TCTTCACGCTTGGTGGGTATCCCGATTTGTGATGACATATACAAACCTTTCGCGG
92092 92118
TGCGTTGTTTTCAGAGTTGCCACCATCGTCTGGCTGGTGGCTGCTGTGATTCAA
92152
TTAAGTAAATTTGGGGTAAGT
...
92361 92411
ACCGTTGGGACTGGCACATGTACGATACCGTGAAGTTCCGACTGGTTCGCGCAAG
92421
ATGCGATTGATTTATGTGCCCGCGCGCCCTGAAAGCCGTATATGA
...
92470 92520
GGAACACATGGGTATGCCCTTTGACCGTGTGGAAGCGTAATAATTTATACCGCTTT
92530
CGCGCGCATACTGCCGAACAGGTAAACCGCGGTAGACCGGCC
...
92646 92696 92702
CGCATTTCTTTGTGATGACGCGACAAGATTTGATTCGTGATGAAAAAGCGGATTCG
92706
TCGGCGTAACCGCCATGGAATGGAACCGCGAAGTTTATATTTTCCACGC
92707
...
92781 92831
GCGGCGCGCGCGTATTATGCGTCTTCTACCAATGCCATATGATGATACCGCGATGGTT
92841
TGGGTAATTTGTCCGCTGACAGGTATCCGTTGGAAGACATGGAAT
...
92932 92982
TACCGAAGCGCTACCGCGAGGCGGTATTCGTGTAATGCCAGCGGCAACGCTTTAT
92992 92994 93032
GGAACGCTATGCCCGACCGTAAGACCTTGCTTCGCGACGTTGTTCCCGCGCAT
|-----|
|-----|
|-----|
```

```

88133 GTTGGGAAGACAGCACTCAAAATCCACACCATATGTCTACTGTGAGTTCACAGCATAACC      |-----|
88193 TGCCGTGGATTGCTGCAATGATGCGGACGATGATCACCATCGAGAC      88183
...
88241 CAGGTTCGACATGGAACCTCTTGACCGCGTTCCGGCAATTCAAATACCCGACGACATCG      |-----|
88301 GCCCGGGGGTTTACGACATCCACAGCCCGCGGTACCGACAGAACGCCAGCTGGAGCACC      88291
88361 TGTTGCCAAGACCATGAGAGTTGTACCGGTTGAGACGTCTGTGGGTTAAACCCGACTTCGC      |-----|
88421 GCCTGAAAACACGCGGCTGGAAAGAACT      88394
...
88695 TTATTTGAACTATTAATTATACAATCATTTTGCATGGGGTAGAATGCCAGCGATTCCAC      |-----|
88755 AATTATTTCTCAAAACCAATCTATTAGAGAGCTCAAAATGGCTTTGC      88745
88801 AAGATCGTACCGGTCAAAAAGTACCTTCGCTAGTATTCCGACCCGCGTGGCGACACTT      |-----|
88861 GGAAGATGTCTACCGATGATTTGTTCAAGGCAAAAAGTAGT      88851
...
88948 CTTTCACACTTCCGCGTTACACGAATTTGCGGCGTTCAAAAGAAACGCGCTTGACG      |-----|
89008 CAATCTACTGCGTATCTGTAAACGATCTTTGTAATGAACGCTTGGGCGCGCGAAGAG      |-----|
89068 AATCCGACAACATCTACATGATTCCTGATGCGAA      89046
...
89111 TACCGAAGGCAATGGTATGTTGTCGTAAGAAGACCTTGCGCTTCGCAAAACGCTCTTG      |-----|
89171 GCGTTACTCCATGCTGTTAACGACGGCGTGTGTAAGAAATGTTCAATCGAGCCGGAAGA      |-----|
89231 ACCGAGCGATCCGTTCAAGATATCCGATGCAAGATCTATGCTGCATTTGTTGCTCCGA      89235
...
89748 TTTCAGACGAAAAGGAGACAAAGATTATGAAAAAATTCAGCGGATGCTGCTGTAA      |-----|
89808 TCGGCGCGGTACTGCCGATATGGTGCATTTCGAATGCCCGTTT      89798
...
89961 ATCAGCATTCATACCGACCGTTGCGGTGCATTTGACAAAGACAGCATGCTGTCA      |-----|

```

```

90021 ACGCGAAGAGTATGACGCGCGTCAAAATCCGAGCGTAGCCGTTTGTCCGCTTTGTC      |-----|
90081 TTGCCGATGTGGAAGAGTGGCCCTGCCGCAAGCCGATCATGTGTTCCGCTAA      90082
...
90199 TGTTACCGGTTCCGGTCCCGTCAATCCGCAATGGCAGTCTTTGGGCAATCGTTGAT      |-----|
90259 TATCAGCATGACGTTTCTCATGTGAGATACGCTGCTAAGCCGTTGCCGTTGCGGCG      |-----|
90319 GGTGT      90263 90269
...
90369 AAGTTGAATTTTCGTTTGGGTGGCATCATGCGCGCATTTCCGACCCCGCTGTTAG      |-----|
90429 ACGAGCGAAAGCCGTTTCGTTGAGTAAGAAATTAACATGCTATTGGA      90419
...
90543 AAGCGTATTTGTTGCCGAATATATGCTGGAGCGGTAGCGCCGCTCCGACGTTGACA      |-----|
90603 ATATCGGTTTGAACACATCAATATCGAAAAAGATGGCGCGCGGCT      90593
...
90654 TTGCCGATCCGCTACTATGACAGACAGTATTCGCATATCTTATTGGCGCGACGCGT      |-----|
90714 CCAACCAACTGCCCTGCTGCATGAGCCGCCGACCAAGGCAMAGATTGCCGCGATTAAG      |-----|
90774 CAGGCGCTTACCCGAAATATCGCGCGGTTTGGCGCGAGCACCAT      90764
...
90869 GCGCAGGTTGCCCGCAATACCAAGCCGACGAAATTTGTCATCGCGAAGTATCGTTAA      |-----|
90929 AACCAAGCGCGACCGCGTGATGCTGTATAACAAGCCATTATGC      90919
...
91336 CAGTTGTTTAAATAACACTGTTCAAAATGGATAAACAGGTAATTAACGTTATG      |-----|
91396 TAACCAAGTAGCAATGGTTTACGGTTTGTAGTCGATPATATA      91386
...
91760 CTGGCGCAATACCGCTAAAGCCGTAATTTGCTTCTGCAATGTTTGTAGCTGTGTTTGG      |-----|

```

```
...
85223 ATTACGAAAGAAATCAATGCGTTAATAACAGCCTTCCGATTTGAAAGCGGACATCA |-----|
85283 ACCTTCGTTGTAATTTTTCACCGCAAAACAGAGCAAAATGCAAAACGATGCGTGGGATT |-----|
85287 85287 85326
85343 CCATCCACCCTGTGCAAAACCTGCAATCCCAAGTTGGTATCCGTAACTACGCGCAAACT
85403 CCGGCAACGCGACCGCAAGCAAGGATGTCACAAAGCATCAACAGAAACCGGCTTGG |-----|
85429
85463 AAGCAGACCCGACCTGACCCGCAATCGACATCCCGCCGACGAATTTGCGCAAAATCGCCA |-----|
85516
85523 AAGACTATTGGACAGCGGCATCCGCCGATTTGTGCGCTGCGTGGCGACGAGCGCCCG |-----|
85555 85567
85583 GTTATGAGAAAAACCGTTTACGCCGAAGACTTGGTTAAGCTATTACGCTCCGCGCGCG |-----|
85636
85643 ACTTCGACATCTCTGTGGCGGCAATATCCGAAGTCATCCGACGACCA
...
85728 AATGATCGGGGTGCAAAACACGATCATCAACCAATTTTCTTTGAGTAGAAGCGGTACCT |-----|
85778
85788 GCGCTTCGCGACCGCTCGTGATGTGGTATGATGATGGAATTCGTCCTCGTATTTT |-----|
85831
85848 GCTGTTACCACTTCAAGCAGCTCGCAAAATGCGCAAGTAAACCAAGTCACAAATCCC |-----|
85894
85908 AAGCTGCTGTGCAAAATGATGAAGTTTGGACGACGACCAAGCGCACGCAACCTCGT |-----|
85939 85942 85966
85968 CGCCGCCAGTATCGGCATCATATGTCAAAAGTCTGTCCCGCAAGGCGTGAAAGATT |-----|
85971 85978
86028 CCACTT
...
86345 CTGTGCTAAAGACTTGGCGGAGAAAAACTGAAAAACACGAGCGCGCTCCACGCGATT |-----|
86395
86405 ACGTTGCCGTAGCGGATTTCACTTTCTAGACACACATCTCGACCTGCAATCGCACCG |-----|
86416 86455
86465 GCGGATTTCCGCGCGCTTGGCTTGCAGCAGCAAAACCTATCTTT
...
86693 AACCGACGTTGTGGTCCGTTGACTTTCCTGTGGGTGCTAAGAAAAAAGCGCGCTTG |-----|
86743
```

```
86753 AATTGACCGCTCTGAGCCTGTGCTTAAACTGTTCCTGTTTACGT
...
86852 AGCCTGCTTTGACTGTTGACTGCTCCCTTAAAGATGGGTAGAACATCAAAAGCTTTACG |-----|
86902
86912 CCACCTTGAGCAAAAGTAAGTGCCAAAATCCTGCTGAGCACTTACTT
...
87032 CTCCTGAACACTGATGATGCTTCGCCGACTAGACAAAGTCTGTCGCCGCGTCATTG |-----|
87082
87092 ACGGCGCAATATTGTGGCGCGCCCAACCTGACAAAGTTTGGAACTGTTGAGCTTTGCG
87152 AAGCCAAACTGGCGACCGCTTTGTGATTTCCAGCTCTTGTGCTGCTGACACTCCGT |-----|
87163
87212 TTGACTT
...
87362 GCCAAGCTCTGCCGACTCCGCTGCCAACAGCAGCAAAATCCACCGTGACGAGCTTGCA |-----|
87412
87422 |-----|
87422 AAGCCTGCGCGATTGCTGCTGCCAACGACGACCAACGCAAAATCCCATTTGGCGACGTA
87482 TCAAAAGCCACACACATGTTGAAACCTGCTCTGCTGCGCGACGACCAATCGTTCTT |-----|
87522
87542 TCCCGCAAACTACGCAAAATCCGCAAGCAGCGCAG
...
87584 AAAAAAGCGAACTGCTGCGCGGATTAAGAGCCGATGAAAAAAGAAATCGCCTTGG |-----|
87634
87644 TGGTTGAAGAGCAAGAAAACTGCACTTGGACGTACTGTACACGCGAAGCCGAGCGTA |-----|
87703
87704 ACGACATGTTGAATACTTCCGCGCAATTGTTGAGCGGTTTGGATTACACCATATA
...
87825 TCCTGAAGCCATGACCGTAGCTTGCTTACTTACGACAAAGCCTGACCAAAAGCGCAT |-----|
87875
87885 GAAAGTATGTTGACCGCGCCTGTAAACATTCTGCAATGTCCTTCTGTCGCAACGATAT |-----|
87937
87945 TCCTCGCTCTACCGTGTGTAACAACATCGCACTGCGCTCTGAACGACGAAGTATGATCT |-----|
87985
88005 GAAAAAGCAGCATCAAAAGTCATCCAAATCGACGA
...

```

```
81913
81965 CTTTCTGCTGCTTGCCTTATTCAGGGGCGGATGCGCGCGGTTGCATAATCGGTTTGGGGC
81979
82025 GCGGAAGACG
...
82483 CCGAGTGGATTGTGGACCGCGGTGGCCAGTAGCCGAATTCGGGAAGCTAGCATTGCGC
82533
82543 TGTTTGCCACGCGCATCAGCAGCGCAGATTTTTCGGTATTCTGTCATCGCTTGGATG
82591
82603 GTCTGGCTCAACACAAATCACAATCAAAAACCTTTGATCGTTGAATGCGGTTAAACCTTCTTCC
82635
82663 AAATCGGCTTGATTAACATTACGCCGCGC
...
82732 CCGTAGCGGCTGCATTTTGTGTTCGACCGAGGCGGCAACGATTCGCGCTGCGCGCAG
82782
82792 CCCAAGTCCAAAGACGCGGCTGCTTCGGGTATCCGGTCAATTCAGTTGCAAAATCATCG
82852 CGCAGGTTCAATTGCTGACATTCCTTATATAAGGTTGTTCAATPAGGCGGCGCAGCCGACGCA
82880
82912 TATAGCCTTGTCTTCATTAAAAAGCATCGTCCCGGTGTG
...
82970 ATACTGCACGGATTTTGGGCGCAATCAGGCGCTTGACCACTTCGCGCAGCTTGGG
83020
83030 CGCGAAGCGCAATCGGTGCTGAAGCTGGCGACAAGAATTTTGCCTGAACATCCTGAC
83035
83090 GCGCGGGGTCAAGCTTTGCCGA
...
83142 GCCTTGGTCATCAGCAATAAGTGTGTGGCATCAAAACCGCCGACGAATTTATTCGCTTGA
83192
83202 TAGGCGAGATAGATTTCTACTCAAAATTCAGCGCATAGCCGATTTGATAGCCGTTGGAA
83231
83262 CGCAATCGCGTCCGAATTTTTCG
...
83356 CGGTGTGGCGGCTGCGTAATGTCCTTCGTTGAATCGGATCGGTCAAAATCGCCTGAC
83406
83416 GTGCTACATCATTAACGCGATATTTGCGTGACAGTTTCGCGCGGACCAATCACTA
83476 AGGATGGCGCAGCGCTCGGATAGGAATTCGTCACTGCAAAAGCCTGCATACGCCCA
83505

83536 AGTGCCTCCGACAAACCGCGGCCA
...
83678 TCCGCTGCTGCCGTCGACAGCCGCCAGATTGTTCAACCGACCAAGAAAAACGTTCTGT
83728
83738 ATCAATCGGTTTGGCGGGTCCATACCATATTTGTCACCAAGCCCGTA
...
83914 AAGTTTACCGTTTCCAAAACCAAGCGCATTTCAACGAATTTTGGGCGCTTACAA
83964
83974 TGCCACCGAGGCAATTTTGAATCATATCTGTTCCAAACAAATGCGGCGAAAACGTTATT
84025
84034 ATATAGCAACGGCATGACTTTTGACACGGTCCGACAAAGCAGCCG
...
84207 TGATTATCAACCGCCTTTCAAGCCGACAGCAAGAAACGGCCCTGCGCGATTCGCGCAAAA
84257
84267 TCAGCGCGTGGGTGCTGCTCGTGACAGCCGCGGACGCTGTTTGTATTGTGTCATGC
84301
84327 TGTATTTCAACACATTCCAGATTGATTT
...
84416 AGGCTTGCGGTAGAGGTGCTTATAGTGATTAATTTAACCAGTACGCGTTGCT
84466
84476 CGCCTTAGCTCAAAAGAGACGATTCTTAAGGTGCTGAAGCAACAGTAATCGGTTCCG
84493
84536 TACTATTTGTACTGTGCTGCGGCTTCGTGCTGATTTTGTTAATCCACTATAC
84561
84596 AACGGAAGCAGGAAGGCGAGGGGTCAAGCTTGGCGTCTTTAAACGCGGATTGCTTTT
84622
84656 GCAGATAGCGTAACCTTTGCCCTGCGGCTTACGATTTGGCAGTCCGCGGTGCGTTT
84686
84716 GCGGTTTTGAGTGAAGACAGAACCTGCATTATTGTCCTTTCTAAGAGATGACATGACG
84761
84776 GATTGAAACGTTGTGAAATTTCTGCGCATGGCTTCGATTAATGACGTTACGCTGTTTG
84783
84836 CCGGTATAGACGAGGATGGATCGGAAGAAATATCCAGCGAAAACAGTGATATCTTTTG
84820
84896 CCGTCTGCCAAACCATCGTTTTCGTTGTTTCGGGCGAAGAACGATACGCGAGCT
84911
84956 TCATTGGCACT
```

```
79254  GTCTGCTGATACCCCTTGCGAGCGCTTGCCTTTTTCGCCGCCAATCCGCCGAGACAA
79256                                79271
79274                                79274
79314  CCCCAGCATCTGATTAAAGCATCATCTTGCCTGCGAAGGAGCGCTTTTGTCAAAATC
                                |-----|
                                79352
                                79355
79374  GTCCCTTTGAGACACCATCAAGCATCATTTGAACAAGATTGATTGAACGTCAAACT
                                |-----|
                                79423
79434  ATGTGCTGTTATTTCGATTATTTTGTCTGATTGTGATTTTGTTC
...
79522  ATACTCCTCCTTTTGTCTGCTGATTTTTCGAAGTCATCGACAATTTGCGGCATATCG
                                |-----|
                                79572
                                79581
79582  -----|
GCGTTGCTGGCGGCTTGCCCGTGTTTTGACCGCGAAGCTCGTTGGCAGGTGCATTTGT
79586
79642  GGACGGACGATGTCCGCCCTTGCGTGCGCTTTCCTTGATTTGCCGATGTCCTTCGCG
                                |-----|
                                79647
79702  T
...
79723  GCACCTTGCAATCCGATCGCGAGATATTGATACCGCGCCTGTTCGCGATGTCGTCATCG
                                |-----|
                                79773
79783  AAACCTTTGCCTGGACCTGCCGAAATGTCTGCATCATCATCG
...
79999  TCCGTTTCGATTAGCGAGCCTTGGCGAAGAGCGTATGCTTCCGGAATAAACGCCCCCG
                                |-----|
                                80049
80059  AATGGCTGCTTTTCGCTATCGGAGCGATGTTTGGCAAAAGTGGTGAATGTGGCGAC
                                |-----|
                                80082
                                80113
80119  AGCGAGGACATCCGTTGACACTTTTGTGGCGGGGCGCAAAATTATCGACAGCCTCAAC
80179  AAAACGGCGTTATTCCGCAAGATGCCCTGCAAAACGCGGAGATGTTTTCAGACGCGCAT
                                |-----|
                                80211
                                80217
80239  CCGTCCGCTCGTCAAAATCCCTTTCGTGCCGA
...
80434  CACACGTTTCTACACGCCCGAAACCGCATCGGACACCGCGCCTTTCAGACGACCTCA
                                |-----|
                                80484
80494  ACGCGGAGAGGCTTTATCCGCAACACACGCGCTCGAATGTGGAATCTCTGCACAC
                                |-----|
                                80521
80554  ATCAAAAAGGCTGGCGCAAGC
...
80707  CAATGAAAAACACACAAGAACTGCCGCCGCCCATGTATTATTATGTCGGCACGATCCTA
                                |-----|
                                80757
80767  TGGTCGTTCAAAAAACCGAATACATCAAAAGCGCGCGCTTTCGCCCAAAAGTCAGCAT
                                |-----|
                                80769
...
80827  AACTGAAAAACCTGCTGACCGCGCGGCTTCGAAACCATTTTACAAAGCCGACGACAAAT
                                |-----|
                                80877
80887  TCGACGTGTCATCCTGTCCCGCAAAAACGTAGCTACAGCTACTT
...
80950  TCTTTATGACGAGAATTCACCAACCAATACGAATCGAAGCCGACACATCGGCGCGGT
                                |-----|
                                81000
81010  TGAATTCATGCTGTGACGCGTATGGAAGACCAATGCGAAGTTTACCTTTCAGCAAGGCAAC
                                |-----|
                                81018
81070  CCAT
...
81085  TGCCACCATCATGTCACGCGAAGTCAGTATACCGAGCTTCCGCTCAAAAGCGATATT
                                |-----|
                                81135
81145  CAGGCAAAAGTATGAAAAACCGCGCGCTGTGTGGGGGACCCGAAATCCAACTGATGCTT
81205  ACATCGAAMACGGCGACMAAATCGAATCGACACCCGCACTGGCGAATTCGCCAAGCGG
                                |-----|
                                81216
81265  CTTGATTTGTTCATTTGAAAAATGCCGCTGSAAAACGTTTCAGACGGCATTTTATCT
                                |-----|
                                81298
81325  TCGCCCCAAGTTTGGATTGAAGTAGATGTTTTTTTCGTAACGCAATACCGCTGATTTT
                                |-----|
                                81364
81385  GCCATTTTCTGCAAAATGATGTCTAAAAACGGTT
...
81597  TCGGCTCCTGCGCTTGTGCCGGGGTTTCTTGACTGCCCTGTATCAGCAACGACGCTGCTT
                                |-----|
                                81647
                                81650
81657  CTCGGAGCGCGCGTTCCTGTTCATCAAGGCATCTTGCTTGCCTGCGCTGATGCTGCGCT
                                |-----|
                                81666
81717  TTCCG
...
81845  TCCAGTTCGTGATGGAAAGGCAACAGACAGTCCGAGTCAATGTGTTCGGGATTTGC
                                |-----|
                                81895
81905  TGCATGAGATGTCATATGCCGGTGGCGGGTGCAGGTGCGTGCATTTGACGGCGGCGACG
```

```
76524 GGCATTATAGCAATTTCCGGTATCCGCTGAAAGGGTTAAACAGGTAGGCGCATGTAT |----|
76574
76584 TTCACCAACAGGATAAACAAGATGATAGCGCCAGCCGATTTTGAACGCCGCTACCGACG |----|
76629 76640
76644 -|
76645 ATTAAGCCCCCAACAGCGTTACCCAAGCCCGCTTACCTGCTGAACATATTTGC
...
76848 ACATAGTCCCGCASTATGCCGCAAGCGTGTAGTCCGACCGCTCCACAAAACGCCCGCG |----|
76898
76908 CCGTAGATTGTGATAGCCCGGAGTAGCGGAGCAGCCAGTTCGGCAAAACATCAATGCC
76988
76968 AATCCGGGCGAAGGGGGGTAAACGATGCCCGCTACCGGCGCTATCAGGGCGAGGGCG |----|
76988
77028 AGGATGACGAGCAAGG
...
77097 GCGGCATTCTTTCGACGATCCGCCCGCTATGGCGGTGCGGTATGAGCGCTTTGTC |----|
77147
77157 GGCAAACAGGTGACGATGCTCTCGCCGCGCGGTTTGGGTTCCCTGCACATAAT |----|
77169
77217 AACACGTT
...
77287 TCGCATCCGTCAAAGCGGTAGTTTTCATTTCTCTGCTGCTGTCACCAAGCGCTTGATT |----|
77337
77347 TCGGCGTGTGGGCTGACATTCAGTCGCAATGCAAGTGTTGTGTGCCGATGCATGATT |----|
77407 TTGCCGCTGAAACGATGACTGCCCGACGCGTATTTCCGCCGTGCGCGCGGATTTGCT |----|
77418
77467 GCTTGGCGGAGTCTTCGACATGAAGTGTTCATTTCTTCCTGCGGGAAGAGCGGCG |----|
77523
77527 -|
77528 ACGGCGGATGTGTTTTCAATTCGGCAAGCAGCGGCGTTTATGCGCTTGGGACATTTCT
77587 TGGCGGCGCGTGCCTCCACAGCAGCTCGAGTTGCCACAGCGTGTTTTCGTAGAGGTC |----|
77608
77647 AACCCGATTCCTTGAG
...
77686 CCGTTTTCGCCGAGTCTTCGAGCGTAGCGATACCGAGCCTGTGCAAGGGCGGACGCTT |----|
77736
77746 TTGGGGCGGAGCGCGCGGTGTACAGCATAGTTTATACCCGAAAA
...
77838 AAACGGTCCGCCAACGCGAGTCTAAAGGTGTTTTCCGCAACCAAGTCAGTTGGCTTGGC |----|
77888
77898 AAGTTGGCGCGGTGGCGCAATGCCCTTTTTCGATGAAGTTTGTCTG
...
78040 TTTTCCACTTCGCCGATTAAGCCCTGTTCAAGCATCAGGTGGAACCAAGCAGGCGCATTTT |----|
78090
78100 TCATGCAAGCGCGCGCGTTTTCGGAATCAGGCGCGGTATGCA
...
78196 GGTTCGCCGTCAGATTAATAACTTCCAAAGCGGCTCCGATACGCTGCTGTGCTCGGT |----|
78246
78256 TTCAGACGGCATGCTGTTTGGGATCGACCTGTTGCAGGGTGGGT
...
78325 TGTTTTGTTCGTCCAGTGGCAGCAGCAGGCGCATCGGCTTCGGGCAATGCTTCAA |----|
78375
78385 CTTTGGGTCAAGGGCGCGAAATACATCATGTCGCGCCGCAATCA
...
78670 TTGCTCCGGCGGTGCGCCGAGCAGGCAAGGCTTTGGGGTGGGATACGTTTCAG |----|
78720
78730 GTTTGAAAATACGATATATAGCGGAAGCGTCCGACGTTATTTTGTGTTGGGAA |----|
78766
78790 GCACGCCGACGCAAGGGGCGGTTCACGT
...
78902 TGAAGCACCAAGTAATCGGTCCGTACTATCTGTACTGTGCGGCTTGCTGCTTGT |----|
78952
78962 CTTGATTTTGTTAATCCACTATAAATTTCAAAACGACGCGCGG
...
79009 TTTTCATATATCCCGTGCCTGTCGCCCTTGTCTCAGGCATCGCGGCGGATCCGAT |----|
79059
79069 TTTTCGGGAGACGCCCGCTTCGCGCGTATTTAAAGTTTCGCGGTGCGGTGTTTCTGC |----|
79078
79129 GGCAA
...
79134 GGCCTTCAGACGCAATCCGCTGCTGCCCTTAGACAAAGCGTGCCTTGGGCGGATPANG |----|
79184 79193
79194 GCGTTTGTCTTTTGAAGCCTTGCAATGTCCGAAACCTGCTTGCCTGCTGCTTGC |----|
79198 79251
```

```
73791
73831 AGGTTGGCGGATTCCTCTGTCTGCCGACGCGACGAGATCCGTCCTGCCGTTTTCATATG
73860
73863
73891 TCCACGTGATTAAGTCGGGAGTTGTT
...
74030 TTGTAGCCTACTTTACCGCAGGTGTAAATCCGGGGGTTGGGATGTGGCGCTTTGCG
74080
74090 CCGAAAGGTATGTGGAATGATTTTTCGTGTTGAATCATTTT
...
74347 GATAGCCGCCGCTTCGTTAATGCGCTCTTGAGCGCGTTACCCATCTTTTGCGGATTTT
74397
74407 GTGCCGATGATTACGCCCGCGCTGATGTGGGCGCGCGCTGGAAATGCCGCTGAGATG
74467 ACGGCGTATCTGCTGTCGATGCGATGCGTGGGTGTGCGACATTATTTGCAAGTTC
74487
74527 AACCGCTTCGGGCGCG
...
74605 CTCGCGCGGGGAGTAAAGAGGCGGTTTGACTAAGATGCGATGATTTTCACGCGCTTGG
74655
74665 GCGGTATGTTTGTGCGCGCGCTTTTGTGTGTTTTCGCGCGCTTCGCGGATTTG
74670
74725 A
...
74821 GCGCCAAAGGCGGACGCGACGTCGCTCGATGAGAAACTTGGGCGTGGCATTCGCTGTG
74871
74881 CTGCTGATTTGTGCTGTGTTCAATTGCATGAATAACCGCTGCTGCGGATGACGCGCATT
74941 GCGGTGCTGTGATTCGCGGCTATATCGTCGCGCTGTTTGTGGCAAGGTGATTTTTCG
74965
75001 GCACGTGCAAAACCTGCCGCT
...
75071 CATTTATTTGTCGGGTGCGATTTTCTTTGTGAGCGTGTTTGAGGCGGTGCGGATTTGA
75121
75131 CGGCGACGGCAATGTTGTCCGACCGAGTATGAAGCGAGAAATACACCAACGCTTGC
75182
75191 GCGCGCGGTGTGTGGGACGCGCTTGTGTGCTGATTTGCGACGCGCTTTGGGTTCCGCTGC
75251 CGCTGACGAGTTTGCACAAACAGCGGTGATTGAGATGACGCGGCTTCGCGCC

73791
73831
73860
73863
73891
...
74030
74080
74090
...
74347
74397
74407
74467
74487
74527
...
74605
74655
74665
74670
74725
...
74821
74871
74881
74941
74965
75001
...
75071
75121
75131
75182
75191
75251

75255
75258
75311
75311 ATG
...
75331 GTGATTTTGTGCTGTGTGGTCTGTGCCCGTGTGTGCGACGCGCGTTTACGACGATTCGG
75381
75391 AGTCGGGTGTGGCGCGCGATGTTTGAATGTTCCGCTTGATTG
...
75451 CCGATTTTGTACGCCACGCGCATTCGCAGAGCGCGGAAGCGGTATTTCCGCAACGTCGGTC
75501
75511 GGTTCGCTTGGGTGTGTCGCTTTGAGCGGAGTGTTTAAAMACC
...
75762 AATACCGCTCGCGCGGCTTCACTTAATACCAATTCCTGCGACACGAGACGTTCAATG
75812
75822 CTTTCGACCTGCGCGACATGGGTATGACACACGACATGTGCAATGTTCCGCTGTACAG
75882 CCGCGCATACCGTGCCTTCTGTTGCCCATCAACCCATGCGCGCTGTGCGGAGATCGCA
75893
75942 TGTAAAGGTGCGGCTTCGCCGCCATATCGGTGCCGATATCCAAATGCCGATTTCTTTC
75953
76002 AGCTCGC
...
76017 CCGCGGAGGTTGTTACGCTGATGAGGAGCGGTTTCGCGCGCGCTGCGGATGATGAGG
76067
76077 CTGACGCTGCGGCTACGCCCGCGCTTTTGTCTTTCGTTGCGATGACGCGCTGATACGCC
76085
76137 ATTG
...
76260 GAGGATCGATTAATCCGACGAGCCCTGATGCGCGCGCTTGTCTGATGCGGCTTGAAGG
76310
76320 CGGTGCGATCGCCAAATATACGCGATGTTTCTGTTTGCCTT
...
76371 AACACTTCGCGCGTGCCTGTGATTTGCTTCTTGTGATGATGAGTTCGACGATGAT
76421
76431 TTAGGTTTTCACAAATCGGCGCTTACGCGCTGAAAGCCGTAGATGGCTTTGTTT
76451
76491 GCCATTAATGTGCTTT
...
```

```
70699 AGCTAAGAGACGATTCCTAAGTCTGAGACACGAGCTGAATCGTTCCGACTAT
      |-----|
      70710
70759 CTGTACTGTCTGGCGCTTCGTCGCTTGTCTGATTTTGTTAATCACTAATACGGCAT
      |-----|
      70778
70819 TTTTCGGCATTTGCC
      |-----|
      ...
71040 CGGGGGCTTCGCTGAATTCGGGGTTTTTCAATCCCGANGTGTCCGATCGTCATCA
      |-----|
      71090
71100 TCGACCAAGCGGCTTCGGGACGTTCCGCCCGCTATGCTTGTGCGGACACAAATACGACTT
      |-----|
      71118
71160 GGGATTTGTGGCG
      |-----|
      ...
71198 CTGGGTAATCGGGAATGCTGTGTGCGCGGTTCTGCGGCGAGTACTTGTCCGTGCT
      |-----|
      71248
71258 TATGCCCAACCCATCCGGAACGGGTAAGGATTGTGTTCGCG
      |-----|
      ...
71568 CAATCGCGCGCTTTGGAACCAATATTTTGTCAAGCGCGCTGTGTGACGGCGCATAGG
      |-----|
      71618 71626
71628 CGATTTTGAAACAATATCGGCACAAATACAGCATATCCGACCATTATCTGTCAGGGGCGGT
      |-----|
      71631
71688 ATGATTTGTGACCCGATGAGTGCAGTGCCTGTGCGAAGCCTTTCCGAAAGCG
      |-----|
      71719
71748 AATTGAGGNGTTCAGGCAAGGCATC
      |-----|
      ...
71933 AATGCGGATGCTCTTTGCGGCATATGCGCTGCCGTCTTTGTGAGCAGGACGATGGC
      |-----|
      71983
71993 GAGCGTTCGGACAGGCTCGGCATAGACAATCAATTCGCGCTGTGCGAGGTTTCTCC
      |-----|
      72001
72053 GCTTTCCTTTGCCACAGGCGCTTTGAACAGCGCGGTTTTGTTCGTAATCTGATG
      |-----|
      72072
72113 CTTTCGTTCCGGGCTTTTGAACCAAGAACCGTGGGTTGCGTTTTGACGACGACG
      |-----|
      72147 72166
72173 GTCAGCCGATGCGGTGAGGGCGAGCGCGTGCGCCGACGTTTCTGCGCTAGTCGCA
      |-----|
      72226
72233 AAGCAATCAAGGCTTTTGCCTTGATACGGCCATTTCGTTGGCGGCGGGAAGGATAGT
      |-----|
      ...
72293 TTTTTCGATGCGTTTTCGCTGCTGTCGCAAGCGCCAAATATTGCAATAAGCGC
      |-----|
      72328

72353 GTCAGAAAGCGGCTTCGAGGTTGGATCG
      |-----|
      ...
72419 TTGCCGCCGTACACTTCTTTCATGCTTTGTGGCGAAGAGATTCGGAACCGCTT
      |-----|
      72469
72479 TTGCCCTGTTCGATACGACGATGAATTTGTCCGCTTCGCCGTGAGTACATGCCGCC
      |-----|
      72496
72539 AAACGACTGTGT
      |-----|
      ...
72673 TTTTCAGNAGCGCCAGATTTCGGCATGAGACTTCCCTCGACAAAGACCGCTGCT
      |-----|
      72723
72733 GCCGTCCGCTCGAGCGGACACCTTTGACCTTTTCAATACTTCGGCATCGCAGCTG
      |-----|
      72760
72793 TTGACGCGCGGTGTGGCGGCTTTTCCAAATTCGCTGGCGCGGACGGCGCCGAACCGGC
      |-----|
      72810 72823 72838
72853 AGCAGCGCGTAGAGTTGCTTGGCTTGGCGGCTTCAAT
      |-----|
      ...
73231 GCGCGGGGCTGGGTTGCGAGAACATGTGTCTTAATCGAATACGACCGCTGCTTAGT
      |-----|
      73281
73291 TCGGTCGAAGAGGATATCGCCTTGAAGCGCGCGGCGCACATATCGGCAAGAGCTTG
      |-----|
      73302 73308
73351 GCGCGGGGCTGGGCTACGCTAATCACGCCCTGTCCGCCGACAGCATGAACGGTAGG
      |-----|
      73356
73411 GCGGTGTGTGCTGCGCGGACAGGACGACGAAGCCTTCGGCGCGGCTTGATGATTCG
      |-----|
      73419 73434
73471 ATGTTGCTGCCGATGTTGCCGTGCTTCCTTCAAGCCGACGATGTTGGGATTTCCGCA
      |-----|
      73509
73531 AGCGCAGAGATGTTTCGTTGTCATGCTGACGACGAGTTTCGCCGGAACGTTGAGAT
      |-----|
      73560 73578
73591 ATCATCGGAATCCGGGCGCTTCGGCAAGCCTTTGAATGGCGGTACATCCCTTCTTC
      |-----|
      73671
73651 GAAAGTTTGTGTAAATAGGGACGACGAGAGGGGTGATCCGCCCGGCTTTTTCGCG
      |-----|
      73671
73711 GCTTGGAAAGGGGATGCTTCGACGCTGTTTTCCTCCCTGTCGCGGAGATGACGGG
      |-----|
      73722 73758
73771 ACGGCTTTGGCAACGTGTTGACGAGCGCTTCATGACGGCGGTGTCTTCTTGACGAG
```


67956 CAGGCAATTTAATTCAAAGCAGCAGCTATTTGGGGATGAGTTTCTGTAACGGGACGA
67988 68011

68016 TCACGGGAAATTTTGCCCGATGATGATCGAGATTGGAGCATCGCCCAAAACCGGTATG
68016 68057

68076 CTTCCGACGAGCATTTACAGATATTCAAGGTTTGAAGAAAGATTATGGAGCAGCGCGCC
68123

68136 GCAAGATGACTTACTCTGCCGCTTTCGATGCGTCCGAAC TGAA
...

68378 AAGAAAAACCCGCATACGAGTATGAAATTTATACCTTTTCAGCGCGCGCGCTTGACG
68428 68434

-
-
68438 ATGAGGATTTTCCCAATGTCGACGTTTACGCATTGAAACGGGCTTCCCTTC
68439

...

68508 AGCGGCTGATAGCCCTGTTTACCCAATCCGGCANTCCGATTTTGACATTGGACGATGAAG
68558

68568 ATTATCGCATAGCAAAATCAAAATGAAAAATGGCGGAGTAAGTAAG
...

68785 GTTTCGCGTAGCGGCTTTGATTTGATCGTCCAGTTTGCCGTTTGCGCGCGCGCTTGAGT
68835

68845 TTTTCGACGGCGCGCGCAATTTGTTTTTCGGCTTCAAAATTTTG
...

68930 CTTGCGGATCGACATCAAAAAAGTGAAAGGTTCTTGCTGCTTTTCGGCAACGCGCTG
68980

68990 CCGCCTGTTGAATTTTTCGACATCGTAAGGTTCTTGCGCTTTGAC
...

69211 CGTGTGATGATGTTGTGTCATTTTCAGCGCGCGCATGCGCTGAGAGGCTTATGA
69261

69271 TACCTGAACAGGTTTGGGAAACGAAAGAAAGAAAAACATGCTGTTTAAATTACC
69331 GGTGCTTCGGCAGATTCGGCGAAGCGATGTGCGCTGCGTTTTCGGGGCGGATACCG
69354

69391 GTTATCGATCGCGCGCGCC
...

69449 CTTTGTTTTACCTTTGGAATGAGACGTGCGGACCGGAGTCGCTGGAAAAACGCTTAA
69499

ACGGCATCCCGGATGAATTTTCGCAATCGACTCCTCATCAACA
69499

...
69562 CTGGCTTTGGGCTCGACACGCGGACACAGCGGATTTTGAAGANTTGGGAACGATGATT |-----| 69612
69562 CTGGCTTTGGGCTCGACACGCGGACACAGCGGATTTTGAAGANTTGGGAACGATGATT |-----| 69612
69622 CAACCAATGTTTTGGGTTGACGCTTCGACCGCCAAATTTTGCCTGCAATGTTGAA |-----| 69642
69682 CGCGCGCGCGGTTATGTGATGAATTTGGGTTGCTTGCAGGCAATTAATGCTTATGCGGC |-----| 69745 69756 69765
69742 AGCAACGTTTACGGGGGCGACCAAGGGCTTTTGGCGCCAGTTCAAGCTGAATTTGGCGCG |-----| 69745 69756 69765
69802 GAGTTGGCGGATAGAACAA |-----| 69842
...
69836 TCGAGCCGGGTTTGTGCGGCAATACGAGTTTTCACACGTGCGTTCAAAAGCGCATGACG |-----| 69886
69896 AGAAGCAGCAGCAGGCGTGTATGAGGGTGTGGAATTTATCCGCCCCGA |-----| 69936
...
70048 ATAAAAAAGCCGTCGCCGAGTGCAGGAAGACTTTGAAAAACAGAGTATGTGCGCTGTT |-----| 70098
70108 TCCCGCATCAGATCTTGTTCAATGATGCGGAATGCCGTCTGAG |-----| 70158
...
70213 ATGGGGTTGCTTCTTTCCGAATCGGGCGGATTTGCCTTCCGCATGATGATGCGACATAA |-----| 70263
70273 GTGAACACGCGCTTCGGTTACGAGTGGCGGCTTCGGTACGCAATCGTTATCAAAATT |-----| 70317
70333 TTCACCCAGACTTCGACTTTAAAGCTGGAGGAAAGTTGTGCCACGCGGACGCATGCCCG |-----| 70387
70393 TAGCAGCAGCAGCGATGTGCGGACCTTGACCGGGCGGATGAAAGTTCATTTCCCTGAACGGC |-----| 70449
70401 70404
70453 ACGGTAAAGATGGGTCCCGCGCGGATTTCCGCCCAATATGCGCGGCCCAAAATCCATC |-----| 70503
70454 70463
70513 T |-----| 70563
...
70519 ATATATCCAGCCGCCGAAATGCTTGTGGATTGATATGCGCGCGCATGACGACGTA |-----| 70569
70579 CGCAGAGCAGTTCCGCTTGAAGGGGTTGGCGGTTGCTTCTGTGCTGATAAAGTTT |-----| 70631
70639 CCTTGTATTATGAAATATATATAGTGATTAACAAAACCACTACGCGCTTACCTCGCTT |-----| 70683

```
65032 TGAGTAGAATAGGCTCCGCGTTGTGAGTTGCGCGGCTTTACGTTTAGCGCATTTGT |-----| |
65082 65090
65092 CGTTCAACAGCAGTTTTCGGTGGCTGCATCCAAAGTGTGCGGAGACTTGGGCTTTGG ---|
65095
65152 CATTGCCGCTTTTTCGCCAATGTTCAAAGTACGGGAGTGCAGGAATTGCCCA |-----|
65181
65212 GCGAGTCCAGCGCAATGTTTTTC |-----|
65181
...
65265 GCAGAACCGCCCGCGGTTTCAACATACCGCGCGTTTCATCAATGGAACGATAGAC |-----|
65315
65325 AGCATTTTTCGCGCTTGTGCCGAGTCCAAATTTGGAAACAAGTCGG |-----|
65315
...
65378 GTCCGCCAAGACATTAACCGGTTACGAGATGTGTCTTCGCGCTTTTTCAGACGACCCAA |-----|
65428
65438 GCTGAACCTTGGCGGCTTCTTCAGAGAGCGAGACGGGATGTCGAG |-----|
65428
...
65580 TTGCTCAGACGGGTGGGTTGACGGCAAGTTGTACCACTTTAACCGGAGGCTTTTG |-----|
65630
65640 GTTTGGACATACCGCAGATGTGCGCGCTTCAACGTCGTGTCATTTAGACTTTTCT |-----|
65660 65672
65700 GCCGCATCAATGACTTGGACTTGGCCGT |-----|
65700
...
65742 AAGTTTTTGTGTGAGCGGTAATCTTCCGCGCTTGCACATCACTCCGAGCTTGGC |-----|
65792
65802 ACGGTACCCATGCTGTCGGATCGAATGCGCGTGTTCGCGCAGAAATCGATGTGGC |-----|
65814
65862 TGATAACGCGCTCGTAGCTGTGTGCGGAATCAGGCTTTGTGTCTGTGCTTTGCCG |-----|
65867
65922 T |-----|
65922
...
65925 TTGTCCACATACGCCGAGATTACGAATCATCGACAGCATGGAAGCATCAACGATAC |-----|
65975 65984
65985 TCCTAGGAACGTGAGGTTGTAATGCTTTATCGGAATCGAACCATTTGCCAATCAGG |-----|
65989
66045 TTGGCGCGCTTAACGCGCAGCATTTTCGGCTTTCGACGCGCGGCTGTCCGCATCCAGT |-----|
66062 |-----|
66062
```

```
60812 ATACGACAGGCAATCCGGGGAAGTTATGTCGATCCGCTTTGGTTGCCGATGTCG 60794
60872 CAACCTTGTCAAAGCTGCGGAAGACATAGTTTC 60850
...
60976 TTGAATTGCTGTAACGGTATTATTATGATTAACAAAATCAGACAGCAGCAAG 61026
61036 CCGAAGACAGTACAGATAGTACGGCAAGCGGCAAGCGCTACTGCTTAAATTTAAT 61071
61096 CCACATACTTCACCTTTTAATCGGCTTGCCC
...
61352 TCTTAAAGCGCAGCGCAATGCAGATGGAACACGCGTTTACGCCGAAACGCCAATC 61402
61412 AGCGGCTCTTATGTTCAACGGCGGATGTTACGAAGTATGATGCCAAACATTGAATA 61432
61472 CCGATCAATAACATAA
...
62268 GGGTGTAAACCTTCTGTGAATCAGGGTAAACATAGGGGGTGTATGTCGCTGT 62318
62328 GTGTTGAACAACATCTTTCGGAACGTGGCAGAGGCTGATGCAAC
...
62635 CGGCAGCGGCGCACATTAACTGTTGGCGCTTCAACAGGTTCAACACATCCGTTTCA 62685
62695 GGTGGCTTTGGCAGCTCACTTTAATCAGTCCGAATAGGCTGCCCG
...
62923 AGCCTTTGTCGGCATAGACGSGTCGTACCTTTGGGAGTCCCTTCCAAACAGCGACAGT 62973
62983 GTGTCACTCATGGGCATTGGTGGAGATGATGACAGTTCTCGAT
...
63054 TGTGTTTGTAAACGAGTTGTAGAGCGCTTTTCTTTATCAACGCGGCTGCTGCC 63104
63114 TTACTCGGTGTGTTGACCGCTGATTTGCTTCTTCTGTCACCTTATAGGCTTGACGC 63169
63174 TGTTCCTGCGCGGCTGTGATATGATGTCATCAACGAGCGGCGGAGATGCTTCTCT 63209
63234 ATTTTAAAGCGCTTTTTCGTCAGTTGCGAGTTAATCAATTGAGTAATTCGACAGGCTG 63293
```

```
63294 TCGTCTTGCGCCAGCCAGTTGGCGTAGCGGATTAAGTGTCTGTAATCGGGATTC 63298
...
63683 AATTTGCAAAAGTCTCGACCTTGTTGTTTTTAAGTATTCGATAGTATGGCGATACCT 63733
63743 TTGGGGTTGTTGTTTTCGTTTGGTTTGGTTTGAACAAAGACGAACCGGATGCAAGTTT 63787
63803 CGTTGGCGATGTCGATATAGTAATTAACAAAATCAGACAGCAGCGAAGCGCAG 63847
63863 ACAGTACAGATAGTACGGAACCGATTCATTGTCCTTCAGCACCTTAGAGATGCTCT 63915
63923 CTTGCACTAAGCGCAGACAGCGGTCACGGSTTTTGTTCATCCACTATAACAGCAACC 63942
63983 CTGTGCGCGTCATTCGCCAANAAGAGGAATCCAGTCCGTCAGTTTGGTCATTTCC 63985
...
64116 CGATTAATTTCTGAGGCATTTGAATTTCCAGATTTCCCGCTGGCGGGAATGCAATTCAT 64166
64176 AAGTTTCCGAATTTCCACATTAACCGAACCCTGACGTAACCGTA
...
64537 GCTGCAGATGCCCAAGGCTTTTATAGTATTAACAAAATCAGGACACGCGACGAG 64587
64597 CCGCAGACATACAGATAGTACGGAACGATTCACCTTGTCCTCAGCACCTTAGAGAT 64655
64657 CGTTCCTTTAGCTAAGCGCAGGCAAGCCGCTACTGTTTTTATATCAATGAGTCCG 64660
64717 GCGTTAATCATTAATCAGGACGATAGGATTAATGCGCTGTAACGAGAAATGTTCA 64721
64777 GACGCAATTTTACCTTTGTCCTTATAAAGCGTTTAGTGCCCTGATTAAGSTTACGCTCG 64805
64837 GACGCATCACTTGTGGGCTTTTT
...
64875 TAGTAGCCGCCGATGTGCGCCGCTTTGCTTTGTAAGCGGAAGCTCGCAACGATTTTC 64925
64935 GCTTCGTGCGCGGTCAAAAGCGGCTGCCAATGCGTAAATGCGGCTT
...
63293
```

```
56752
56762 AATCTTATGTAGCAATATATGATGATGATATATAAATGAAACAAAAGGGTTTACATTAATGAT
56822 GAGATGATGATAGTCGTCGGCATCTCGGCATTTATACAGCTGTCATTTGCCATTAACCTTCTTAT
56882 CAAAGTTATATTGAAAAAGGCTATCAGTCCCA
...
56928 TGGTCGGTATCAACAATATTCCAAACAGTTTATTTGAAAAATCCCTGGAGCATATC
56988 AGACCATCAAGACCAAACTGGAATATTGTCTCAGGCTATTAACAT
...
57063 TTTGCGTGCAATTTGTCAATAGAGAAAAACCAAGGCATACAGCTTGGTCGGCTTCCAA
57123 AGACGGGACGGGTTATACTTTGTGCGTATGATGAACAGCGTGGCGCAGATACAAAT
57183 GCCGTGATGCCGCTTCTGCCCCGAGCCCATTTGGAGACCTT
...
57301 TTTTCAGACGATTTTATGGGTATAGTGATTAACAATAATCAAGACAAAGCCGCAAAAG
57361 CCGCAGACAGTACAATAATAGTACGGAACCGATTCATTGTGCTTCAGCACCTTAAGAAAT
57421 CGTTCTCTTTGAGCTAAGGCGAAGCGCGTACTGCTGTTTGTATATCCACTATACAT
57481 CCCGTCATTCGCCAAGAAAGTG
...
58430 CGATTAATTTGCTTAGCATTTGATGATGCTAGATTCGCCCTGCGGGAATGACGATTCAT
58490 ATAGTGGATTACAAAAATCAGACAAAGCGCGCAGCTGCAGACAGTACAGATTAATAG
58550 GAACCGATTACCTTGTGTCTCAGCACCTTAGAGAAATCGTCTCTTTGAGCTAAGCGGAG
58610 GCAACGCTGACTGCTTTTGTATATCCACTATAAAAAGCATATTGAATGCGGGCAAAC
58670 CGGCTGCTTCCGTTTTTGGATTTGAGAAATGCCATCGCCAGCTTTCAACACATTA
58730 AAAACAGTGGGCGACGCTTTTTCAGCGG
...
58783 AGTGCCGCCCCCATCAGGATATGCCGAGAAATTAATCATAAAGTTACGGTGGCGATAAGC
58833
58843 AGTATGGCAGAGGTTCCGCCACAGGTTGACCGTGGCAAACTCCG
...
59149 CGGCTTGTATTAATCTCCCATCTTTGAAACCGCGCGTATGACAGAGCAAGACGATGAAT
59209 ATTGAAGTAGAATGAAGTATGAGCAGACGATGGCGGATGTG
...
59267 CAACGAAGGTTCTGCAGGTTTAGATTTGGCGGCTGTTTGSATGAGAAAGTCGTTTGC
59327 AGCCGGTGAACGTTCTTGTGTGCGCAGACAGTTTGGCAATTTATTTGGCTGATCCCTCAT
59387 ATGCCG
...
59734 AATACCTTACTCAACACAGCTCAAAACCTATCCCTTTGCCCGCAGCTCAGAGGCGATGACAG
59794 GGCATTTCCGCGCCTGAAGGATGGAAGCCGTCCCTGCACATCG
...
59916 GCTGACTGCGGCTCTGCTGAATGCTGTCAGGCGTGTCAAACTGTTAAACGCGCTTA
59976 CGATGCGTTGACAGTGGATGGCGGATTAATGAATTTCTGCCGTTTTAGCAGATAGGAGGCG
60036 GTTGTCTTCTTTGTTCAAAACCGTGTGAACCTGTTTCAGACGCGCATCAAC
60038
...
60446 TCACCAGTTTGTCCAAGCGGTCCAATGTTTCCAGGCTGCTTCCGTTTTTGTGCGCGGCG
60506 ATGCGGAATGCTTAAAAATTTTCTGCTTTACAGAACTATACGGCAGCGC
60507
...
60632 TGTATCAGGAAAAATTTGAGGCGGTTATCCCATTTTGCACACAGTATTGACGTTAAAT
60692 TACCGGATGCTCGTTTACATCTGTTGAAGAATCCCGATGGCGACGATTTGGCATTTG
60733
60736
60752 CACGCAATTTATGGCAAAAAAGCCGCTATCAAGTATTGCCCGGACGCTTTTGGCAGCGG
```

53160 53198
53220 TGATGGCGGGATCCGGCGCTCCGACACCGTGTGAGAGCTTGGCGAATTTCACGTT
53280 CCCTCAAGACGTTGGCGGAAAGATTGCAAGTCCCATCTCCCTGTGCGCAATTGAGCC
53287 53318
53340 GTACCGTGAATCCGGTACCGGACAAACCGCCATGATGTCGACCTCCGGCAATCCGGTG
53364
53400 CTATCGAGCAAGATGCCGAC
53594 ATAGAGATTAAATGGCTATATATAAATTTATTAGGCGAATCAGGCAAAATCGTTTAA
53654 ATCATGCTTAGAGATTGCCCTAAAAATTAACGGCGTCTTGAGGC
53710 TCAGCCCGCATATAATGAAATATAGTGAATTACAAAATCAGACAGAGCGGCAAG
53760
53770 CCGGACAGTACAAATAGTACGGAACGATTCACCTGGTGTGTCAGACACTTAGAGAA
53830 ---|
53833 53855
53890 TTACCACTTAACGTTCTATGCTTATGTGTACACGAAACAAACAGTTTTCACGCTAAC
53899
53950 GAGCTGCTCAGTGTGATGGTCAATGACCGCTTATGGGAGATGATAGCCCTCCCAATATG
53985
54010 AGCCAATGATTCATCCCGCGCATTTGCCA
54149 TCACAATATAATGTGACTCCGCAAGAGGGGAGGAATGTGGCTTTTGGCGATAGA
54209 ATGGCAATAGACATATGACGCGCATGGCGGATGTTTCTCCGACAGCGTGTGTTGA
54229
54269 ATGATACGAGCAGACGCCGATTAATTATGCTTCAACCAATATGCGTTGCGTCTC
54277 54310
54329 AGCCGACGCGGACCGTGTACTTTGACATTCATCAAAAGGAGCGTTCGTTATACGA
54372
54389 AAGCCAGCATCTTACAGCAAAATCCAGCTTTTATT
54497 CGGGCGTGTGTTTATTAACAGCAGCGGCGGTTGAGGTTTGTCTAGAAACGATAGCC
54547
54557 GTACCGTATGCCAATATAATGATTGTTTATGTCGTTTGGGAAT

54606 AAAATGAGAAATATGATTGCTTCCGCCCTGAAAAACCCCAAGTCCGGTATGGCGGTGATA
54666 ---|
54675
54726 TTAGCGACAGTGCCTTCCGTCAGGGAGCGAGACGCAAAACCATGCTGATCAAAATCAG
54735 54768
54786 CAAAACCTGATGGAAGAAATGTTGATGAATCCGACAT
54844 AAAACTATTAATCTTACATGGGAAACATCATGCACTATCAGTTGTGATGCGATTTTC
54904 AGTTGATGCCATAAAACTAAGACCGAGTTGGCAGAGGCACATTT
55096 ATACTTTGATTAACTATTGTGGTAATGATTGGCAGGGGATTCGGATATCGGCCGTA
55156 CGAATCTTGAGACGAAACGGCAACAATATGATTAATACCTATCAGGC
55330 GGTGGATCGAGTTACTTACATCCCGGAAATTAATGATGCGGCAAGCGCTTTTC
55390 CGCGCACAGGATTTGCGGAATGCGCAACAATGATGTTGCGCGAT
55965 TTGACATATGAAGTGGGTATATCTATGTTTCCGACTGTCTGAAGATGACGATGCGG
56025 GCAAAAGAGAAAAATTCAAATATACGGGTACATTGACAGCTCCAC
56231 CTCTGACGACAGAGGGGTTTGCACTGTTTATGCTGTGATGATGATGATGATGATG
56291 CTTTTTGTGTTAAGTCCGCGAGTCTTACAAATACGAGACAGC
56520 GGCCTTTGACAATATCGTGTGCAAGGCAACCCAGCTTGAGGCGGTGAAGCGTTCTTG
56580 CACTGCAAAATCTACAGGCTGTGCATTTGACAAATAAGAGATGAATATAGAAAGGAC
56640 GCAAGCGTCAAGCAAAATGCCAGCTTATATTATGCAATATTGGCGGTGAAGAACGAGA
56644
56702 ATGTTTATCGGGTACTGCGCAGGCTTGCGGTAAAGATGCCAATACCGTGTGCTCTTC

```
49234 TGTGCTATGTTGGTAACGGCGAGTTTGACGAAGTAGAGAGTTTGACGGCGATGACG          49284
- |-----|
49294 ACGAATGGCTCGAAGCGCACAGGTTGCGGGCACCGCGCGCTTGCGGCTGCGGAC          49332
49295          49332 49335
49354 ATTAAGGCATTGCAGGCTTGCCGCAAGGGGGCAAG
...
49393 TTTGCCGTTTTTAATAATATCGTGATTCATCTGTTTGCAATGGGCTGAACGCTCC          49443
49453 TTTGATTTTATGTAATAAAAAACTATCTTGATTATATATTA
...
49965 GGGGAGATTTTGGACGTTAAATCTATATTTTACCAAAATCAATTAATTAATGTTG          50015
50025 GTTGTTAACAACATTCGCGCACCCCAATTATTTTAAACATCATCCGAACTGATTGG          50086
50085 CATGCCCTTATATATGTCAAAGCGGTGATTGTTTATGATGCGCTGCCCATTTATTTGCT
50145 GGTCTTATTTAAGCGTATCCGACAGAAAGAACGGCGATGGCAACATTTTTAACCGCT          50178
50205 TTCAATCAATTAATAGCGGATTAAACAAAACAGTACGGCTTGCCCTGCGCTTGCGTACTG          50240
50265 GTTTTGTTAATCCGCTATATTTCCGCCATCT
...
50484 TTTGATATATGCTGTTGTTCTAAGTCTAGATTCGCCGCTGCGGGGATACGATTCAT          50534
50544 AAGTTTCCGAAATTCACAAATTAACGAAACCTGACAGTAAACGTA
...
50675 CGCAAAATCTACCGGAATGATATAGTGAATTAACAAAAATCAGACAAAGCGACGAG          50725
50735 CCGCAGACAGTACAGATAGTACGAACCGATTCACTGCTGCTTACGACCTTAGAGATC          50792
50795 GTTCCTTTGAGCTAAGCGAGACAGCGCTACTGTTTTTTGTTAATCCGCTATATTC          50819
50855 GCCATCTCTAAGATTTCAG
...
51217 TTTGAATATTTGCTGTTGTTTAAGGTTAGATTCCCGGCTGCGGGGAGATGACGATTCAT          51267
51277 AAGTTTCCGAAATTCACAACTAACGAACTGACAGTAAACGTA
```

```
...
51416 TCTACCCGAATATATATAGCGATTATACAAAATATCAGACAAGCGCGCAACGTTTGGC          51466 51473
-- |-----|
51476 GACTTCGTCCAGTTGACGATTTCCAGAAAACCTTTAGATGATTGGGACGGCTGTGCG          51478 51491
51536 GTAGTCGATGT
...
51561 CACAGCTGCAGGTACGACGCGGCTGTTTCAGTGTGCAGCGGCTAGCGGCTTGGAA          51611
51621 GTGAAACCAAAATCCAATCCGCGGACAGGGGTTTTCACCAACCAAG
...
52152 CGCTATATAGAGGTTGCAATTTTCGCGACGACAGATAGATGTACATGAACGATTACA          52202
52212 CAGCTATGCCCTCTGAAGACCGGTGAGTGCGCGCATTTGCTGCTCCACATTCGATGG          52249
52272 AGCGGAAACAATCCGTTTGGGCGGTTGATGC
...
52510 AAACACCCCGTCTGCGCCCAACATCCCGCTAGCGCGAAATCGTGCGCGAGCGTTCCAT          52560
52570 TATGCGCACTCGCCGAAGTGGGAGCGGAATCGCCCGCGCTTAACAATCCGCAAG          52601 52610
52630 CAGGACCGCGGGGCGAGCTTTTGGACGAGCGGAAACAATATTCAAATCCGCGAAG          52679
52690 CACGCGCAAAATCCAAGAGGGCTTTTGGAGATGCCGATTTTGCT
...
52800 TGTGACAGGGTTCATCGACCTCGACAAAAAAACCTCGGTCGTCAACCCGCGACCTGA          52850
52860 TTATCGTTGCCGTCGTCCTGCTATAGGTTAAGACCGCCTTTTCCATCAATATCCGCGAC          52862 52872 52910
52920 ACGTTGCCGTAGAAGCAAGCTGCCGTTGCCGCTTCTCAATGAGATGSGGGGCGC
52980 AACTGTCATGCTATGCTCGGCTCGGTGAGCGCTTGATCAAAAGGTTTTTGAACCG          53010 53024
53040 GCAGGCTCGAAGACGAACACTGGGGTGTGGAAGCAAGCATGCTCAAACTCTCGAG          53064 53081
53100 CGCCGTGTACATGACGAGACCCCGGGGTCTGACTGCGCTCGAACTGCGCGCCGCC          53160
53160 GTGCTTCGCCCGCTCAATTTAACAATAAGTGGGATTAAATGTCATGACCTGCAAC
```

```
46741 GTGCCAAGAGTGCCAAATACGCCGGAGAGGGCGCAATGCGCGGTGAAGTT 46734
...
46842 CGTATGTGTCCGATACCAACGCTGACCGCGCGGAACCTCGCCCATCGCGGGCGGTGGTG 46892
...
46902 AGGATGATGCCGTAGAGTAACGCCCATTTGATGTTGGCGAGGTTAACGGCCAAAAACATC 46920
46962 TGCCAGCCGCTTGCGCCGAGTATCAATGCCCGCTTCTTCGCTGCGCCGTGACCTGC 47006
47022 ATCAGCGGGATGATTTTCGCGCGCGCGCAAAAGGGAAGTTAACGACAGCGTCCCGCAAAATA 47067
47082 ATGCCGGGAGTGGCGAAGATATCTGTATGCCCTTGCGCTTCGAG 47070
...
47137 ATGCCGTGTGGCGCGCGCAACAATAAGACGAACATTAAACGGCGCACCAACGCGCATACGG 47187
47197 AAAACGCAAAATCGAGCAGGGTAGTCAGCAACTGCTTGCCTCGGAATCAAAACGGGTCA
47257 GCAGCAGCGCATCCCGCACACCACATACAGCATTTGACGGGAACGACATCAGCGGGTAA 47297
47317 TCAGGTTAAATTTGATGSCAGACCAACGCTTGCGGATTCGTTTAAGGATTTTAAGTACAAAT 47351
47377 CCCAACCGCCTTTTAAAGGCTTCGTAAACACGCGGAGGCGGCAACGACGATCAGCA 47408
47437 GCAGAAAGCCACGCGCGGCAATCA
...
47510 ATAGGTTTCATGGCGGTGTGTCTCTGTTGTTTAAATGTTTGGAGTTCGTGAA 47560
47570 AAGGATTTAGGGGTTTCAGACGACCTTTTGTTCATTAACTAGTAGGTCGATACCTTGT 47589
47630 ATCCGACAAGGTGC
...
47669 CAAGTATCCGACCTACGAATTCGTGTTTGGGAGCGGATTCGTTTTCAGACGACCTTT 47719
47729 TATCTTTTATGCTAAAGAGAAATAAAAAAGCTAGCTGCGCTCAT
...
47840 TTCACCAATGGATTCCGCGCTGCGTGGGAATGACGTAACAGTATTTTCAGACGCCGCGCT 47890

47900 CAACCTTCGCGCGCGGAAACGCGCTTGCCCAACGCCCACTGCATGCATTACGCCAACAACA 47916
47960 ATCACAACGAACACGACGACATTAACAACAGCCACCGCGGACGCCCTGTAGCTGAAC 47987
48020 TGTTCAGCTTCCCGTATATAT
...
48218 GCTTCCTCATATTCGCCCGGACAAATTTCTCCATATACCGGCTGCACGCGCGGCGGATTAAG 48268
48278 GGCAGGCTGACGACGACGACGCAATCCAAATGCCGACGGGTGTAAACGCAATTTTGATG 48286
48338 CCCAAAGGCTGAAAAAAGGCTTATCCAAACCGTTGGCGGCATTCAGAGGCTGCAGAACGG 48386
48398 ATACCCGTAAACCGCGCTGCGGACGCGCAAAAGGCAAAATCGAACG
...
48483 ATACCCAGCGCACGAGCTGCCGAACACGACATTTGTCAGATCCGATAAACGACATCC 48533
48543 GCAAACTCAGCCAGACGCGCGCCAGACATGGTTCCGTAATCCT
...
48609 CGATTTCCGCGCGCTTCCGCCCATCATCGCAACGGCAAGCACAGCAGCGCAGAGC 48659
48669 ACAATACGCTGACGACCAAGGCTGAGTTTGAAGCCGCGCAGTACGCCGCGTTTGAAGC 48716
48729 CTAACATAAACATATGCTGAATAATAGGAAAAAGAGACTTACTTTAACGATGCCGTCG 48730
48789 AAAAAAGGAAAGAAATGAAAGTTGTGTCAAAAGACGAATTTG 48775
...
48856 AGCGGGCGCGATGTTTAAATAATATGTGATTAATTAATTAATCAAGATAGGCGCAGAG 48906
48916 CCGCAGACAGTACAAATAGTACGGCAGGCGCAGGCAAGCGCGTACTGTTTAAATTAAT 48951
48976 CCACTATACGCGCTTGTGTTGACGAAAAACCATCATATATAAGAACTTATGAGATTT 48983
49036 TAT
...
...
|-----|
|-----|
```

```
44308
44355 AAGCGTGGCGGAGATGTTTTCGGCGTGTTCATATCAACGATGCTTAAGATTATGC |-----|
44395
44415 GATGAGCTGGCGGAACAGAGGGCTGACCTATATCGCACCGTTTGACGATCCTGATGT |-----|
44461
44475 GATTGCAGACAGGAGGAGTAGGATGGAATTGTCAGCCA
...
44573 CGGCGTGGCGGCATTATCAAGCAGTCCGTCCCGAATCAAGTTATCGCGTTTCAGA |-----|
44623
44633 |-----| CCAACGATTCTCTGTATGAAGCAGTGGTCGAAGCGGCTGMAATCGTCCATTGGAAG |-----|
44635 44677
44693 ATGTGGGCTGTTTCAAGCGGACGGCGGTCAAACTGCTGGAAACGAAACCTTCGTC |-----|
44728
44753 TCTGCMAAGAACTTTGGATGAATCATTTAC
...
44815 AAGATATTTTCGATGACACGCGCAGCATTTACGAGCCGGCGGCGGCTTGCGG |-----|
44865
44875 GGTGGAAGCCTATATGCCCCGAGAGGGCGGAAACCAACCTGATTGCCGTTTACC |-----|
44890
44935 AGCGGTGCGAA
...
45059 TCAATATATTTGGAAATAGAAATATTACGAGTTCACTACCGCTACGGGACGATGAAA |-----|
45109
45119 AAGGCGCATCTTGTGTGGACTTCAAGGGGCAAGGCCCGCAGAGATT
...
45237 AATCCATATCCGCTATATGTCGAGGGCGGACGACAAAGTAGAAAAACGAGCGTGTG |-----|
45287
45297 CAGTTTGAAGTTTCCGAGCGTCCGGGGCGGTGGCGGTTTTTGAACCATATGACGGG |-----|
45314 45324
45357 CGGTTGAACATTAAGCTCTTCC
...
45483 CAGCTATACAGAGAAACGCAAAATGCCGCTACAGCTGTTTCTTGCTGCTGACGCTTGAA |-----|
45533
45543 AGCACAATGCCGCTGAAGCCTTTCAGACGCAATGGCGTTTCAT
...
45619 TGAGAAAAAGACATAAACCTGTTTGAATTAACTTAACTTGCCTGCGGCAATGGG |-----|
45669
```

```
45679 TAACCGCGGATCGCTCGCACGCCAGGTGATATGACAGTCTTGTGTTGCCGTTTCACC |-----|
45690
45739 AAAATAT
...
45763 GCGTGACTTTTTCGATTTTCGCACAAATCATCGGTGTTTGTGTTGGCGGATCTGC |-----|
45813
45823 CATTGTCGGCGGATATAGCCGATGCGGTTGTTCTGTCAT
...
45889 CACGCGAAGCCGTTATATATGCCAGAAAGCCTTTTCGATGCTCCTCAAAAGCGTGGTT |-----|
45939
45949 TCGCCGAGAACTCGGTACGAAAGCGTTTCGGGTTTGGGTAATGCTTCGGCACTG |-----|
45972
46009 CCGGTTGTGATTTTCCGCTGTTTCATCACGACGATTTGTCGAAACTTCAGAGGCT |-----|
46041
46069 TCTTCTTGTCGTGCTAACCAATGC
...
46177 GGTTCGTCMAAAGCAGAGATTGCGTTTCGACCCGCAAGCGCAGGCGGAGGATGCGC |-----|
46227
46237 TGCGGTTGCCCGCGGAGAGTTGTGCGGATAGGATTTTGGCAAAATGAGAGCTGCACG |-----|
46238
46297 AGCTTGACCAATTCCTGACTTTGCGCGGATTTGTCTTTGACGGCGCTTCGACTTG |-----|
46343
46357 GGCAATACGTCAAACGGAAGCAAGCTGTGCMAACAGTTCAATATGCGGAAAAAGGCG |-----|
46380
46417 TAGTGTGAAACGAAAC
...
46453 ACGTGTGGCGGTTTACGCTTGCCCATCAAAAGAACTTGGCCGCGCTGCTTTC |-----|
46503
46513 AGTCCGCGATATGCGTAAAGCGTGTGTTTGGCCGACCGGACGGGCGGACAGGAA |-----|
46518
46573 ACGAGTTTGC CGGTGGGAGGTTGAGGTGATGTTTTTCAAGCGGTGAATAATTGCCAAG |-----|
46590
46633 TGTTTGTTAAGT
...
46681 CGGCGGAGTTTTTGTCTTGTATATTTGTGATGATGTTCTGCACCGCGGCTGCCCA |-----|
46731
```



```
...
41115 ATCTTCAGCCGACCTGTGCGCCGACCAATGACAACTGATGGAATCTCTGACGATGCGG |-----|
41175 GATGCACTGAAGCGTGTCTCGGACAGGTGCTATTACCACAGCATTC |-----|
...
41362 GTTCTTCGAAATTCGGTGACAAATATTATGCCACCCCGATTCGTGTGAACGACATCA |-----|
41422 AAGCAACGCGGATTGAAAATCTGACCGTCTAGCCCGGACATCGCGGCTGTCTCCGCG |-----|
41482 CCGCGCGGTGGCAAAATCTTGAAATGCGGACCTTGGCAATCATGATTAAGCGCCCGCA |-----|
41542 AAGCCAAATGTGCGAAGTCATGAAACATCATGCGGATATTCAAGGACAGAACCTGTCTGA |-----|
41602 TTGTGAGAGATATGATTGACACTGCATAATACGCTGTGCAAAAGCGCGCTGCCTGAAAG |-----|
41662 AGCGGGGGGCGAAGCTGTCTAGCCTATGCGCAGCCAGCTGTATTCTCCGAGAGGCGG |-----|
41722 TCAGCCGTATGCGCTCATCGAAATGCAACAGTGTGCTGAACCGATACCATTCCTTTGT |-----|
41782 CTGAAGGGGCTAAMAACTGCGACCGTATCCGTCAGGTACGATTCGGGCTCTGTTGCGCG |-----|
41842 AAACCGTCCGCGCATTAAGCAATGAAGAAATCCG |-----|
...
41914 TGTGTGTCGCATAGCCCGAAGCCGCTCTTAAGCTGTGCGCGCGGATGACGCGCATTTTAA |-----|
41974 CCTAAGTTGAGATATTAACTGACTTATGAATTTCAAGCTTCTGT |-----|
...
42064 AAGGCCAAATCCCGGCAATTCGTACGGTGAAGGTCAAGAGCCGTTGCAATCCCTGTGG |-----|
42124 ATCACAAAACCGTATTTCTAGCCATTGSAAAAAGAAATCTTTCATCTACTGCGTTGATTAAGT |-----|
42184 TGTCTCTGAACGTTGAACCAAGACGTTATCGTCCGCGATTTCCAAATGACACCCGTTCC |-----|
42244 GCGCGCAAGTTCAACACATGACTTCCAAAGCTGTGAAGACAGACCAACCCGTACGATCC |-----|
42304 GCGTCCCTTGACACATGCTTAACGCTGAATAATTCACAGGGGTCAAACTGCAGAGCGGCC |-----|
42364 GCGTATCTCTGTTGAA |-----|
...
42406 CTGCCAACATCCCTGCTTTCTTGATTTGGATTGTGCCAAGTGGTGGCGGCAATTC |-----|
42466 TGCACCTGTCAAGACATCAAACTGCTGAAGGTGTAGAAAGCGTTTCCGTGAACGTAACG |-----|
42526 AAAATCTGCGCTGTCTACCGCTTACCGGTGAAGAACGCTTAATTTGATTTTCAGCAGCAGGC |-----|
42586 GGTGCGTATGCAATACGTACCGCCCTGTT |-----|
...
42707 CAGGCCAAAGCCTTTGCGACCGGCTCTTTTATACATTTTCCAGTGGCGGATTTCT |-----|
42767 TTTTGGCAATGTGTATCCGTTTGTCTGATTTTGAATTTTCTTA |-----|
...
43324 TTCTCTACTCAAGCCTGTGCGATTTTGTGAATACAGTGTCTTCAATGCCCAAGCGTGGCG |-----|
43384 TTCTTTGTTCATTTGTTGTCACAAATTTTCAATGACGCGTTGCCAGCGCGCGGCAAG |-----|
43444 GGTAGGCGCGCATGCTTTGGGATAGTCATATGCTTTTAAAGTTGTGTGCTCT |-----|
43504 GACGTAATCGCCGGGACGTACAAACATTCGCTTCTCTGAAAGCCATGCGGATTCGGG |-----|
43564 TA |-----|
...
43661 GGGGATTTGATGTTTTGGCGGATTAATCTGTTTCTTTGAAGTGCATACGATGTGT |-----|
43721 GCGGCTGTAGAGGTTTGGGTTGTTTGAACGTGGGGCGCGTGTGCGGTCTGTGTG |-----|
43781 GCGCGGGCGCATGACAGCGTTGCGTGAGAAACGCTTAAGATGATGAAAGCAGACGGGC |-----|
43841 AGGATTTTATGTGCTGTATGAAT |-----|
...
44032 AATTACTCATCCGATCTGACGGCATCTGTCTATGATGTGGGGTGCAGAACGCTTTTG |-----|
44092 GAACGCGCAGCAGCCTTTCTGTACGTTTGAAAAAACAACATCTTTT |-----|
...
44235 GATTGCGCAAGCGCGGCAATCAACGCAAGGCGTGCATTGTCCGACAGCGTTTGGG |-----|
44295 TTGCGGTGCGGTATGCTCATGCGGAGACTACGCCGCAAAATCAAAAGTGAATGCGGTCAA |-----|
```

```
...
37185 GCGCGATGCGAAGCGCCGGAAGAAAGTCGAAAGTTTTCAGAAAGCAGCGCTTACAG |-----|
37245 CGAGGAAGAAATCAAAAAAGAACGCGACGGCTTCCGCGACGTGGGC 37235
...
37358 CGGCTTGGAACCTATATGCTGATGTTGGAACGACAAATCCCGCGAGTCCGCCAGAC |-----|
37418 GCGCCTTGGAATGGCCGTGCGCTGAAAGCGGTTTGAACAGCGGAATGATTTATCAGA |-----|
37478 AATGGCGGAGATTG 37437
...
37648 GTGCAACAGACGGGTTGGCGCAAAAAGCATCGAAAGCGTTCCCGCGCGTTGAGA |-----|
37708 TATGACATCTGCCGGAAGCGGCTTGCCGATGTGTTCAGCGTACAGGAGCGCA 37698
37768 AAGAAAAAGCAATCGAGCTTTGACAGGTTTGGCGAAGCTCGATACGGAATATTGCC |-----|
37828 CCCACTTTAATGAGTGTGCGCTGCTGCTGACAGCAATATCCGAATATCTGACGGCTTT 37839
37888 TTCGAGC
...
37987 TTGAACGTGCTGTTGGAACGCAATCCGAAATGACAGACCTGTATTTATTCAGCGCATATTG |-----|
38047 GCGCAAAACGAAAAAGAGTGTTCGGTTATCGACGGCTACGCCG 38037
...
38098 GCATACGGCAGGGGAGCGGGGAACAGCGGGGAGGGCGCAATGACGGCGCGCATGATA |-----|
38158 TATGCCAGCCGAAGGATTACACCAAGTCAAGCAGTGGTTGAAAA 38148
...
38494 GTTTACGATGGCTTGGCAAGCGAAAAAATGATTTCAATCTTGAAGGGCGCTTCAGG |-----|
38554 CTTGACCCGATTAACGCTCAGATTATGAATATCTGGGCTACAGCCTGCTTTCCGATTCC |-----|
38614 AAACGTTTGAAGCAAGGCTTCCCTGCTTCAGACGCAATACCAATACCCGAGCAT |-----|
38615 38615 38668
38674 ACCGCTTCAACGACACATAGGCTGGGCGTATTACCTGAAGGCGGGAAGCGCG |-----|
38683 38716
```

```
38734 CTGCCGATCTGCGGATTTGTTGAAAAAGACCCCGAGCCCGAAGTTGCCGCCATTTG 38761
38794 GCGGAAGTGTGTGGGCAATTGGG
...
38839 GTATGACGACAGCGGCAACACTTACGGGAGACAGAAATATATGGCGGGAACGTTCAAA |-----|
38899 38889 38897
---|
CGTCAGCGCATCGCATTGCCCCAACCTTCCGAAAACTCGGAATATATGACAG 38902
...
39509 CAATACGGTTGAGCTGTGCGGAGAACCGCCGACAGTGGGGGCAAGTCCGAACGTTGCNA |-----|
39569 CTGAATACGGAATTTGACATCAGGCTGTTTCACCGAGATTG 39559
...
39650 GCGGCACGCATACAGTAAGTAGATATATGAGTGTACGAGCAGCAGCGAGGCTTCCC |-----|
39710 GCACCTGCAAAACGAAATCTCGATTGAGGATTACCGGCAAGCGGG 39700
...
39780 GCATATTTGCTGCTGATAGATTTCAGAGTACCGTATATTGAACCAAGGAGCGAGCA |-----|
39840 AAATCATCTGTCACAAATCCGGTTGATGGAATGCCAGAGGCGGATTAACTACCGTG 39830
39900 CCGATGCTTGTGTCGAAAAATATGACGACGCCCTTACCGGCTGAGATATGGCTGAGCA |-----|
39904
...
40092 CGGACGTACCGTTTTTTTATTTTCGCAAAATGCGTTGCGGGGATATAGCGCAGGC |-----|
40152 TGAAGAAATGATATTTCCGAACAGTGTATGTCATCGTCAAAACCGCCGTCACGTTT |-----|
40212 CCACCGCAAAATTTTCAACACGAGAGCTT 40187
...
40486 AGAGGCTTTCAAAACATCTTTGTGTGCTATTAACATTGTTGGGAGTCGTCAAGC |-----|
40546 GGTAAAGACACTGATTTTGAATCCAGCATGGAAGTTGCAATCTTCTCCCAAGCA 40536
40606 AGCCAAACGAGTTGGGAGTCGTCAAGCGGTTAAGACACTGATTTTGAATTCACAGATGC |-----|
40666 GAAGTTGAAATCCTTCCCTCCCAAGCAGCCCTTTTGGGAGTGTCTCAAGGCGTTAAGA 40708
40726 CACTGATTTTGAATTCAGCATGCGAAGTTGCAATCC
```

```
32844 CAAACAAACGGTATGAAACGTAATGAGCGCTTGCCGCCAGCAAGACGATTGGCGG |-----|
32904 TATCCGCTCGATGTGGTTCCGCTT                                     32873
...
33321 TTTTATCCGTTGAGGATACAGGGCCGGAAGCGAGATGTGTCCGCCAGACGATGAGT |-----|
33381 TTGCGGTGAAGAAATGCGTGGACCGCGATGTAAATGACGGCAATCTTGCGGTGCC |-----|
33441 GCCCATTTTCAGTGTCTTCCCGATTTGCGACGACTTGCTCTTCGCGCCGCTGTATCATACAGC |-----|
33501 GTTTGGCAACATTGGAAACGGCGGACGCGCTGGGGCCGCGGTATAGTGCA |-----|
33561 GCGCGGATGAGCAGCAATAATCGGGTATGCGCGCTTGTCGCGCAATGTGCGACATPA |-----|
33621 CCGCCGAAGAGAACCGCGATPAAGCAAAATTCGGGGCTTCGGAT |-----|
...
34003 CCGTGAAGATAGTGTCTTAAAGGTGGAATGCTTCCGTTGTGGGTTGGACGCTTCA |-----|
34063 GGCCGCAAGCATACGGTGAAGTCCATCCGTATGCTGTCTCAG |-----|
...
34176 GCGACATTCGCGTGGCTTTCGSCAGAGATTCAAGACGGCCCTTCAAAACGCTCAAC |-----|
34236 AACCGTTCGATGAAGCGGTGTGAGCGTTGTCCGCGCTCCACCAAGTTGGTTACGAATATG |-----|
34296 CCGTCTGAAGAGAGTGGCTTTCGGC |-----|
...
34446 ATTCAAAATTTCTCGCCCTCGAAAGGCAACTCAAAACAGGTTGCGGGCAATGCGCATGAC |-----|
34506 TGAAGATTGATGTCACGCGCGGTTTGGCGCGTGTCCGGCAGATAGG |-----|
...
34656 GAGAGTACCACTTCGAGCGGCTGTCAGAGTTCAATGAGCTTTGAACGGTCTCCCTGCC |-----|
34716 AAGTGAAGCAAGCATATTGCCCTTCTTCGGAATGCCGACTTCGG |-----|
...
35191 TGATGATTACTGTGACACACTGCMAAAATGAAGCGCGCGCGAGAAATATCGCTATGCG |-----|
35241
```

```
35251 TGACCGCTTACGAATCCAGTTTGGCCGCTGATGAGCATGCGCGCTGATGTCTCGC |-----|
35311 TGTGTGGGATTCCTTTGGGAATGGCGTTCAAGGGCGGCAATGTCAGATGCC |-----|
35371 TCGCGCATATGTGTATATCACACCGAATGTGTAGCACGCGGTGCMAAAATGCGATGATG |-----|
35431 TCAGGATTTTCCCGCTTTGTGTGATATCAGACAGATGAAGAGAGAGCGGTTTGGCCGCCG |-----|
35491 CCGAATGATGCTGCGCGCGCGCATATGATCAAACTCGAAGCGGTTGTGTGATGGCGG |-----|
35551 AAGCAGCCGAATTTCTGCAAAATCCGCGGATTTCCGGTCTGTGCCACATCGGTTGA |-----|
...
35619 GTGTTTGCTTGGCGGATACAAAGTTCAAGGGCGCGCGGCAAGCGCGGCTTGCTT |-----|
35679 AAGCATGCCAAGCGCGCATGATGATCGGGGGCGGCTGTGTATGAGTGTCT |-----|
...
36192 TGCACGAAGGGCATCTTGCCCTCTGTTTCGGAAGCCAAAGCCGCCGACAACTGTG |-----|
36252 TCAGCATATTCGTCAACCGCTGCAATTCGGTCAAGGCGAGAGATT |-----|
...
36355 TGCCGTTCTTTTCGCGCCCGATGAGAAAGAACTTATCCGAACGTGAACAGCGCTTACAA |-----|
36415 CGTCGAACCGCCCAATCTGCAAAATGAGTTGTGGCGCAAGTTTGTCCGGACATTTCTT |-----|
36475 CGGGTGGCGACGGGTGTGTCCAAAGCTGTCAATATGTGCGCGCGATGTTGCCCTGTTT |-----|
36535 CG |-----|
...
36851 GCGCGGGGCGGAGATPAAGAACTGTGTCTTGGCCGCCCGCTGTGGGACAGCGCC |-----|
36911 CTGATTGACAAATTGGAATPAACCTCCCTTAAACCGCAAGCGTGGGAATGCCGTCTGA |-----|
36971 AGCGAATTTGCGTTTCAGACGGCATTTATTTTGTGAACGGGTTTGCMAATCTACAAA |-----|
37031 CGATTTCGCTTGTGATAAAGTTACGCTGATATATGCGCGTGAAGTTCCGA |-----|
37034
```

```
...
27911 ATATGATATGATCATCATTTGCTTATCTTATGTTTTCACATTTGGGCGGTAAAA |-----|
27961
27971 GACTGATTTAATGCTGCACCATGCAMATATGTAATTAAACAAAAACGATACGGCGTTGC |-----|
28023
28031 TTGGCCTTAGCTCAAAAGAACGATTTCTTAAGTGCTGAAGCAACGAGTAATGCGTTC |-----|
28050
28091 CGTACTATCTGTACTGTCTGGCGCTTGGCGCTTGTCTGATTTTGTATATCTACTAG |-----|
28118
28151 AGTTAAACAACCTTATTTTGA
...
28238 TGTATGTATGATTTAGACTAAAGAAATCATTTGCTCAATTTTAAAGCGTCTCCC |-----|
28288
28298 CGACAGGTTTCCGCTCAGAGATCAAACTCAAGAAAAACGTATTA
...
28628 TTCTCTCGATTTGGCTTGTATATTTTGTAGGGTTACGATAGTGTATATCCCTTTAT |-----|
28678
28688 TTGAAGTCCACACACACAGAGAAATAGCCTATTTAAACATG
...
28869 GAACACTTGGCTTCAAAATGCTGATGTGTCCAAAGCCGTTGCCATTTGGGCGGCAATATA |-----|
28919
28929 GACAGCATCATCGTGGCTCAGTGATTTGTTTTTTCATACCAAAAT
...
29333 AGATGTTGATATATCAGATAAATATATATGATTAACAAAAACAGTACAGCGTGCCT |-----|
29383
29393 TGCCTTAGCTCAAGAGAGATTTCTTAAGTGCTGAAGCAGCGAGTGAATGCGTCCG |-----|
29410
29453 TACTATTTGTACTGTCTGCGGCTTCGTGCGCTTGTCTGATTTAAATTTACTCCACTATA |-----|
29478
29513 CAAAAACATCAAGAACTATT
...
29546 AATCTGAATTTCTTTTTTTGATGAATTAATAATGAATATATGTAAGCTTTCAA |-----|
29596
29606 CGAAGCGCTTAGCCTCCGATTCGAGTATTTCAATTCAGATPAA
...
30031 ATTACTGTTTTGCCATTATTTATAGTGATTAATTTAAATCAGACAAAGCGCAGAG |-----|
30081
```

```
30091 CCGCAGACAGTACAAATATAGTACGGCAAGGCGGAGGCGGTACTATATCATATATAT |-----|
30126
30151 ATCCATTATATAGAGAAAAATTTCTTATACC
...
30968 GACATTAGCTCTCTTGTATTAAGCTTGTGCTGCAACACGCTGCGTAACCCAGCGTCGC |-----|
31018
31028 AAGSATCTGCCCGCATGCAAAATTCCTATTTGCCACCAACGATGCC
...
31290 TAAACGTATGCTGTACGCCAAATTTCAATTGCGAGACTGTTTGCCCTGAACGTTGAAA |-----|
31340
31350 CGTCGGATGCGTTTAAAGCCGTGTTTTGTGTAACCGAAACCTG
...
31452 CGGTGCGGGCTGCTTGCCCGCATTTGATGATGCCGCCGCGGTGAGGCTGATTTGCTGCGC |-----|
31502
31512 GCGAAGATATATTGGGATTCAAGCATCCAGTAATTTGCTGCTTTGT
...
31569 CTGAAGAGGTTTTGCTGCCGTTGATACGGTAGGACGCGTCAATGAGAGAGACGCGG |-----|
31619
31629 TCGATGGCTTTGTAGTGTGGGCGCATGAGCCAGGATTTTCCGACGAGGCTTTGTTG |-----|
31649
31689 CGGATTTTTCGTAACCGTCTTTCAAGAAAGCTGATTAGGCGGCGGTTTTGTGCTCT |-----|
31742
31749 TTAAAGAGGTGTGCTGATGCCAGGATACGTGCAATCCGTTTGT
...
32035 TAAGGGAAGTTTCCAGTTTCCATTTGCCCTTGTGCGTCATTATGTCTCAATGCTCAAGG |-----|
32085
32095 GCAAGTCCGATTAAGCGGATTAAGGCAAGATAGCGGCGCAAGCGCGG
...
32407 GGCTGCCGCTGTGCATTTTCTGTGATCCGCTTTCTTTGTGGCGCGGACGATTTGGC |-----|
32457
32467 GCAGGCGGTTGAACAGCGGTTTTGTGTCGTTCCGATGTGTGGGCGTTGAGACAACTT |-----|
32510
32527 CTTGTTTGAGCCGGGTAAGGTGATGTTTATTACCGTG
...
32784 GCAGGCGGTATGCTGCTGTGCTGATTAATCTGCAACGCGGAGCCTATACAGACGATGTTT |-----|
32834
```

```
...
22852 CAAATCAGGTTTTTCGTATCCAAAAACGGCAAGAGATTGTTTTGAGAGCGGCTTGGCT |-----|
22912 GTCCTCATCCAAATTATCTTCAATACGGGTACGATATTCCAACT 22902
...
23077 GATATCCGGCCCTTCAAAAACCCAAAGCCTCGATACCGACCTGATGAGACTGACGATTAACG |-----|
23137 GCCTTTTTCGAGCGCTCGCGGGGGAACATCGGCCCATATACACAGCTTTTTCGGGGCT 23127
23197 GTTGTAC
...
23225 GCACGACGAGCAAGAACCGGTACCTTCAGGAGCAAGCTCAAACTCAAGAAATCGTTTGA |-----|
23285 TCGGAGAGGTGTACATTTCCTTCCGCAACATCGGTTTCCTCGC 23275
...
23376 GGATTTGCTGATTAACCGTAAGTGGGTGTCAAGCGCGACCGCTATCTCAAAAGCGCTGCG |-----|
23436 AAAACGACGCCGTCAGTTTGAATCTTTTGTCTGACAGCAGAGAGTGTGTTCAATGCTT |-----|
23496 TGACGATTTGATTTTGTGTCATTTCAAGTAAGATGCTTAATCAAAATGCGGGCGA |-----|
23556 |-----|
23556 TTATAGCGGATTTTAAAGGTTTGTGAGGTTGGGGGTGTTTGGGAGCGC 23551
...
23757 GGGTTTGGCAGCATCCAAAAAAGTGGCCCTGCTCATTTGAACAATGACCCGAGCGGGT |-----|
23817 CGGCAAGCAAAACCGTATTGCGGTCTATCCGCGCAATACGAGAA 23807
...
23940 AGCCTTTCGCTTCAAAAACCAATCTGGCATTAATGCCCGCATTAATCTCAACGACGCGC |-----|
24000 GCATCTGCTCCTTATCCAGCTTTTTCACACTTCTTCTTCCGTACAGGCTTGGCCGTA |-----|
24060 AATTGTTCAAAAGGCTCGCCACGAGAGCGCCCGCAGAAATAATCCAAATCTGCTTAA |-----|
24120 CAATATTGAA 24071
...
24134 CGCCTTCTTCCAACTGTCACTTTGATTTTCCGTAAACAAGATTAATGTTAAAC |-----|
24194 ATCGGTGACGATTCAAACGATTAAGTAAGAAAGACACACAGCGCAACGAAAAACA |-----|
24254 TATTGAACCTTCATCATATATTTCACAT 24210
...
25222 CGCTTCTTCAATTTCCAGTATTTCCCCATTTAGCTTAACGCAAGAAAGAGAGAGCCATTT |-----|
25282 CGGCTTCGTTATGATTAACCCGCTTCGTTGATATAGACTTCGCG 25272
...
25506 AGTACAATAATAGAGGAACCGATTCAGTTGGTGTTCACAGACCTTAGAATCGTCTCT |-----|
25566 TTGACCTAAGCGGAGCAACGCGCTACTGTGTTTAATTTAATCCGCTAATACATCTCAATA |-----|
25626 CGATAAATATTTC 25583
...
26026 TAGTTCAAGATTCATTTTTCGTCAACAAGTAATCCGATTTTGAATTAAGCTTTTCA |-----|
26086 TCAGCTGCCGTTTATCTAAACCGGCAAGTTGAGTTTCAAAATTAAC 26076
...
26392 GAGTTAGAGCGGTTTTTTCATGATCTGTGCTAGATTTGATTTCCCAAGGCGCGATGACG |-----|
26452 ACCAATTCCTGTCCAGGAAAGTCTTGGCCGTATTATCCGTAACCTGACGATGATATG |-----|
26512 GGGAAATTTTCCAAATAGATGACCTGTTCTATTAACCGAAAGG 26499
...
26573 CCGATTTTAAATCTCAGTTCATATTTAGCAAAAGAACAACTTTCCTCGCGCGCTTACA |-----|
26633 CCACCAAGAAATTTGGCCTAATCCGCCAGCAATAGCAACATCTTAACAGAAAGTGTCTG 26623
26693 CCTGTGTTGATTAATAAACAATGCTGTGTCCACATTAACAAATGACGACACCAACAG |-----|
26753 ATAGCCAATGGAAGAAACGCGCCCTGTGTCTCTTCACTCTCTTTGAGAAAGCTCCGCC |-----|
26813 AACTGCATCGG 26768
```

```
19723
19731 CAGCAGAGACTTTTCTCTTTTTCAGGCGGACGCCGCAAGGTAAAGTCGGAGGAGAAA
19759
19791 CGAGTTCGGGGAGAAATGAACCG
...
20066 TCAATACGGTTGCTCAGGCTGACGGCGGCGAGCGGTTTTTTCTTCGTTTGTCTGCGGCG
20116 20122
-|
20126 TTTTCATACGGGTTTTCCGAAACATTTGATTGAATCCGCAACGGCGTCTTCAGGTTG
20127
20186 AAGGCTTTGCGGAAACGTTTGGGTGAGATAGCGCGTATAGCTGTGGAATCGCGTGACAG
20198
20246 GAATTGCC
...
20290 GTGGGCATACGCAATTTTCGGAGCGACCAACCGGACGCGGCGGTTGCTGACGCTCGAT
20340
20350 TGGCGTTTGCAGAGCGCGCTGATTTTCGGCGTCCGATCTTAATCATCGCGCGTTGTA
20378 20396
20410 GGCAGCCTGAATGCTGTCAAAACCAACGCTATACCGCGCTC
...
20482 AAAATCGAGAAATACAGTTTGGGTTGATATCGGTTTCACTTGCTCGCAGCGTTCTTC
20532
20542 GCTGATGCCGTCCTCATTTTATTGACCGCCACCAACAGCAGCGCTGCTCCAAAGCGAA
20602 ACCTGCATTCGTCGCAATCTTGTGCGGCAATGCTGCTGCGGTGCATACCAAAACGAG
20609 20626 20659
|-----|
--|
20662 GACGTTTGCCTTCAACCGCGCTGCAATGCTTGTGATAAGGAGATTTTTCACGTCTTC
20664 20687
20722 ATCCACTTTGCCCGGACGCGG
...
20783 CGCTGGAATGATATGATACTGTCCGGCGTGTACCTGCCATATCGAAGCGGATGACG
20833
20843 CGCTCTTCGCGGAGATGGCGTTTACCAAGTTGATTTTACGACATTTTGACGACCGATA
20860 20892
20903 ACGGCAAAACAGAGATGTTTCGCTCTTCTCTTTTTCAGGC
...
20972 TCAATCAGATATATACAGCCGTCACCCGTGCGCACCTGAATATACATAGGGTCGCCAAA
21022
21032 GCAGTTCTGTAAACTCGCGCGCAAGCAGCGCCCTGTTCGCCCTT
...
21099 AAACGGGCGCGGCTTTGGCCAAACGGTGGCAATGATTTTATCTTGGCGGTCAGAC
21149
21159 CGGTACGGCGCTCCACCAAAAAACACACCGCATCGGCTTCATCGAC
...
21353 ACAGTGCCTCTTGTGTCGCGCTCAACGGTTGAATAAAGTAGATTTGCCGACGTTGGGG
21403
21413 CGGCCGACACGCGGATGTTGTTTCAATGATGATCTTTCTGTG
...
21668 TGTCCGTAGTTTTTTAAGGCTTCCGTGGCTTTTCCCTGTGCGGATAGACATGCTTTTG
21718
21728 GTTTCATCAGACGAGGGGGGCAAGTCCGCTTCAACCGGCGTGTGAGTGGCGCAAGCGCG
21839
21788 GCATCGATTTTTTTTGTGTCACACACACGCGCCAGACGCTGCGCGCCACAGCCTGG
21817 21825
21848 ATCAGGCTGTCCTTTTGTGTCGACAAATCCATTTCAAAATGCTTGGGCAACATCGTAA
21906
21908 CGTGCAGGTCAAATTCGGTTGCTGCGCCATCAGCGTGCTTGGCGGCGGAA
21911
...
22220 GCCATTAATGCGTCTTCCATTTGTTGTAAGTGTGTTGTCAAATCAGCGCGGACGCTT
22270
22280 TGTGATCGTGTGCGCGCTTCATATCTTTGAGCGTAACCGTAACGT
...
22469 ACCTGCAAGTCCGCCCTTGCCTTGGTGATTCATAGACATCAGCGCGGCTTACT
22519
22529 TCCAGGAGCCGTATTGCTCACCAAAAGCAGCGGCTGATGC
...
22629 CGTGTAAAGCGCGCGCCGACACAGTTGCTGCGCGCGAGTTTGTGCTGCTCCACT
22679
22689 CAAAACCGTTGATGTAAATATCAAAACCGCAGAACCAAGCGGATTTTCAATATATTT
22749
22749 GGATACCCAAACGCTCAACATCGCTTTGAAGCTGATAGTGTGTTGGCAAGCTTCGC
22769
22809 CCAATATATCCACGAG
```

```
16018
16021
16024
-|
16028 TCACAGGAAACCAATATCCGTTGCCGTCGAAACAATAGATTCCGACCCG
16029
...
16121 TCGATATAGACCACGCCCAAGGTTTCGGTCCGCGCTGGCAAGACGAATGCGCTGTTC
16171
16181 CTCAATCAATCCAAAGACTGCTTCGGGCTGCAATGCCCCAAAATG
...
16278 CTTCCTGTCCGTCGCCCGCATGTTCAAGCCGAGTTGGCGCATTTGTCGAACGCCGAG
16328
16338 CAAAGCGAGTAGAGACATCGTGTCTTTGAGACGGCGAAGTG
...
16579 CATTGTGGGAGCGGATCAGACCGCGGTGATGGTGTGTGTTCCGTGACGACATAG
16629
16639 GGGATGCCGTATTTTGGAAATTTTGGCGAGTATGCCGCAT
...
17560 TGTTCAGGCAAGATGCCGCTTGAATATACGATTTCCATGTGTCGCCGACGCCGCA
17610
17620 CGAGGTATCGATAGGTCAAGCAAGATGCCGTTTTCGTTTCAAGGCCGTGCCGAAA
17680
17687 GTTGGCGGCGCTTTGGGTGCGTCCGGTTAAATGCTCAAAATTAATTTGGGACACAT
17714
17740 CGGACGCCATATCGCGGAGGGGAGCTGCTCAGGCAAAAATGCTTCGTCCATGCCGAT
17800
17807 GCGGCAAGGCTCAATATGAGGGTGGGATTCGCCCGCGCTGTGAGGAATCAG
17860
17864 CATTAGGTCATGGAAGGACGTTCACTCTGATTTAGAGTGAACGCAATGCCCTGAA
17920
17924 ACCGAGCATTAAGGTTCAAGCGCATTTATTTTCAAGGCTTTGGGAGGGTAAAGCGCG
17972
17980 TTTGCCCATATATTTGCCGTTGGGCTTTTGTATGAGGTTTGCAGATTTCTGCTCT
18032
18040 CGAAGAAGAGACTTGGCCACGCCCTTCGCTGCTAGATTTTGGCGG
...
18102 TTGAAAAATTCAGAGGTACGTACCTTCCATTCGCGTTCGAACGGGTAAAGCTTACG
16018
16021
16024
-|
18162 AATAATGCCGACGGCGGTAGGTGATTTGCCCAAGCAGACGGTCAAGCTTGGCGGG
18163
18222 ATGCGAAATATTGACCGGTGGCGCCAGTGGGAAGAAATTGGCGGATGATGACAGC
18281
18282 TCGTCTCAACGTAACGAAGTTTTCGGTCAAGATTTTGGATTCGACGATGCTG
18286
18342 TTGATGTGTGAAGATTTTAATTCGTTTGGCGCAGCGGATGT
...
18740 TCGAATTTGAACGTTTCGATAAATTTTCGGCGTTGGCGCAACGTAATAACGCTTGT
18790
18800 CCGTCAATTTGCCATTTTAAACGAAACCGGATATGTCGCCCGCTTGGCAGCGCAT
18852
18860 GTGCCTTTGGCGGGGCTTTACCGTCTTTGCCGACATGCCCTTGTGATGACCGGGT
18895
18920 TCTGATAGAAATCAGAACGGCGACGCG
...
19103 CTGACGAAATCCGCCATCAGGATTTGCGAGCGAAGCGGACCGCGCATCGCGCG
19153
19163 TTGCTTTGGGCTTCTTCCGAAATAGCTGAGATTTGATGCCCTTGA
...
19289 TTGACGTGTGAACCACTTGTGCGCGTTAAATCTTTGATTTGATTTGACGACGGA
19339
19349 ACAATATCGCCCGTTCGCTTTGAGTTTGCATTCGCTGTGCAATTCAAAACCGCTTCCAA
19378
19409 TGCCCTTGAACAGCAGCCGCTCATCAGGAAAGAGCGCATTTGTTGTAATTCGAGC
19445
19469 GAGGCTTCAAAACCGCCAAAGTTGGCGCAG
...
19551 GGTAGCCGTGGCGGTGTTAAACACGCTTCTGTAGCGTGTGCGGTTAAACGCTCGGG
19601
19611 GACGAGGTGCTGTGTAATTTGCTTTTAAAGCTTGTGTTTTCGACACGACATTC
19618
19671 GGGCTTCATTTGTTGATTTCCGCTACCTTCTTCCAAATATTCGGGATGACGCCG
```

```
13494 CGGATTTGACATACCGGACGACATCCATCCGCTGCTCAATGCTTTGGCGCTTTCAGACGG      |-----|
      13510
13554 |-----|
      13555 CATGCCCGGAAAAATGACGGTGTGCTTGCCCTGAAAAAATACCGATGAGCGGCTGCGTGC
      ...
13617 ATGCGTTTCGGTGGTTTCGTA CTGCGGCCCTTGATGTGCGCCGAGTTGGACGCCCATG      |-----|
      13667
13677 GAAAAACAATCTGCACAGGTTTGA AACCCTGCCCCCTGTCC
      ...
13725 TGCATCAATACGATGCGCGAAACCGTGAACGGGTTGGGCGCTACCGCAACGCTTTTC      |-----|
      13775
13785 CGCCACGATGCGCGAAATTTGGA TGTTCAGCCCGCAGTTTGA AAC
      ...
13900 TTGCGCGCCGAAACCGGCAATGCGGACCGGCCCTTGGCCGATGTGTGACGCGGCTTTCAG      |-----|
      13950
13960 ACGGCATCGCAGGAACAACGAGCGAGATGACCAAAAAAGAACGCCGCCCGCACAGAA
14020 AAAAAAGAGACGAGGAAAAAAGCGCCGCTGCGACGGCCAAACCCAAAGCCAGAGTT      |-----|
      14042
14080 GTGGGAACGGATTGGGAA
      ...
14289 TGGGCAATTTGACCGCGTGTGCCGCATGCCCTGCATTTGCCGACGAATCGCGGATTTGG      |-----|
      14339
14349 AAGCAGCTCTGGGCGAGTTGGAACACCGCTGTGCGGCTATTGGA AAGGCGACGATACCG      |-----|
      14378
14409 TCAAAATGACCTTTTCGTTCAAAATTCACCATGTATGTATGCAAGCGTACGCCCTTTC      |-----|
      14453
14469 AGAAGACGTTTGAAGGCAAGCGATCGGAATGAAGCGGTGCGCGGACAGAAAAGTGCCTCAGG      |-----|
      14473      14486
14529 CGTGCAAACCGCGA
      ...
14681 TTTACAGCGGCATTTGAATGTCA GTCGTTTTCGCCATTCGTAGAGCTTGAGAGATTTT      |-----|
      14731
14741 TGCACCAAAAGGATGCGGACACAGCTCTTCGCGGTAAAGGTGTGA AATACAGCCCTTCC      |-----|
      14761
14801 ACGCCGTGCAGTTCT
      ...
14822 GCATCTTTTAATCCCGATTTGATGTTTTTTGGCAGTTCGATTTGGCTGTGCGCCGGTA
14882 ATGACGGCTTTCGCGCCGAAAGCCGATCGGGTCAAGAACATTTTCA
      ...
14931 GTTCGGGCGTGTGTTTTTGGCCTTCGTGAGAGATGATATGCGCCGCTTGAGCGTCTGC      |-----|
      14981
14991 CGCGCATATAGCGGAGGGGCAATCTCAATCAGACCTTTTCAAT
      ...
15170 ACTAAAAATGATGCCCTTGCACTTG GTGTTTTCCATCGCATCGACGGCGGCGCGCGG      |-----|
      15220
15230 AGATAGSTTTTGCCCGTACCTGCGCGCCGAGACCGAATACGATGTGTGTTGAGTAAG      |-----|
15290 GCGCGATATAGCCGTTCTGCGCGCTTCTGCGCGCATGCTGCCGCTTGCTGTCGG      |-----|
      15314
15350 AAATATATAGCGGTGTGTCATG
      ...
15377 TCTTGATGACCGGATCTTCGTTTGGCTTCGACGCGGCAAGCCTGATGTGCGCTCG      |-----|
      15427      15433
      -|
15437 TTTAGTCGCGCGTCTGTGCCGTTTCC AAGATTTGAGCAGTGGCGCTTGC
      15438
      ...
15547 ATGAGTGTCTTTAGCAATCAATCAAGGTTGTGTCAAAAGAACCCGACAGAGCTTCAA      |-----|
      15597      15605
      -|
15607 CGCCAACTGTGCGTTTCTTCTAATGCAAGTGGACGATGTGTATGATGAAG
      15610
      ...
15671 TTGGATATGTGTGATTTTAATCTAATAGTGATTAATTAACCACTACAGGTTGCC      |-----|
      15721
15731 CGCCTTAGCTCAAAAGAGAACGATTTCTTAAGTGTGAGACCAACCAAGTAAATGAGTTCCG      |-----|
      15748
15791 TACTATTTGTACTGTCTGGGCTTGTGCGCTTGTCTCTGATTTTGTTAATCCATATAT      |-----|
      15816
15851 TTCACGGATTTTCTTACCG
      ...
15968 TCAAAATCAGAGTGTGTTCATATCTTGTGATGATGCGTGGCAAGGCAATCGTGTGCG      |-----|
      15972
```



```
10724 TAAAGTTCACGCTATGTCCGTCAAAATCAAAACCGGATCCGTAATACCGAGCTTTCA 10774
10784 ATCATTTGCTCACCAGCCGCGGATCCCTTAATCGCCCGCTTTGTGCTCGCGCGGTG 10801
10844 TGCAAAGTCTGTCCGAATTTGACGCAAACTCGCTTCCCTCCTGCTTATCAAAAGCTGA 10864
10904 CGAATTGGCAAGCCGCGCCCGCGCTTTGGCGGATGCGGTTGAGGCTCAGGAAAAATCT 10896
10964 TGATTTGCGCGACATACGATGCCAGCGCGGACCGCTGTGCGCTGCTTTGACGCTT 10969
11024 TGGCGGGGATGGGGGCAAAAGTACATTTCTGTGCCAACCGCTTTGAACACGGCTACG 11028
11084 GTTTAAGCGCCGAGCTGGCCGAATCGCCCGCGCAAGCGGTGATTTGCTGATACGG 11088
11144 TCGATTACGGCATTTGCCAGATGCGAGCGGTGGCGCGCGAGCTGTGGGTTTGATG 11107
11204 TCATCGTTACTGACACACACTTGGCGGCGGAGACCGTCCGACATGATCGTCAATC 11206
11264 CGAACCAGAAAGCTGCGGTTTCCAAAGCAAAAGCTTGGCGGCGTGGCGTGAATTTT 11254
11324 ATGATTGATGGCGTGTGCGGATTTGGCGCCCGCAATTTATTTTTCAGACGGCTAA 11334
11384 AAGAGCCGAAATCTGGGAGCTTTTGGATTTGTGCGACATCGGTACCGTTCCGACGTCG 11398
11444 TCTCCCTCGATACACAAACGTAATCCTCGTGTGCAAGTTTGAACGAT 11445
11555 GCGCAAGGCCACAGCCGTTGATGCGGCTTTGGCGTTGGCGCCGCGCATCAACGCCCGCG 11605
11615 GACGCTGTGACGATGTGCGGTGCGCATCGCTGCCCTGTTGGCGCGAGATGATTCGAG 11623
11675 CTCAGSAACTGGCGGCTCACTTAACAACCTCAATATCGAGCGTCCGGAATCGAGCAGT 11714
11735 CTATGCTGAAGACGCGCTGAATGCTTTTCTGAAACCTGCTTCAGTTCAGATGACTT 11714
11795 TGGTGGCGTATCGGAGCACTTCCATCAAGGCGTGTGCGCATGTTGCCAGCGCCCTCA 11809
11867 AAGACCGTTTTTATGCTCCAGACATCGTGTTTGGCCCTGCGGACAGCGCGCAAGTACCG 11867
11915 GTTGGGAGCGTTTCCATTCCTCAATTTGCACGTGCGGATGCTTGGACTTGTGTCCAAAC 11921
11975 GCGATCCCGATTTGATTTTGAATTCGGGACACAGCGATGGCGGCTTTGAGCATAC 11977
12035 TTGAACACACATTCGCCGCTTTCAGAGACACTTTGAAGAAAGCCGTGCCGAAATGCTGT 12060
12095 GCGAAGACGATTTTGTGCAAACTTCATCACCAGAGCGAGCTGCGCCCTGCGCATCA 12100
12155 C 12100
12185 ACGTTTGGGGGAGGCGCTTCCGCGCCGAGCTTTACCGAGATTCACGCTGCTCGCC 12235
12245 AGCAACCTTTGGCGGCGAGGCGCAAAACAAAAAGTCTGCTCAAAAAAGACGCGTGGC 12235
12305 AATTGAAAGCATGTTTGTGGGCTTGCAGCGAAGACATTTCCGAATACATCCGACAGTTT 12312
12365 ACCGCCCCGTTGCC 12323
12815 CTGCCGAAAGACACACATCCGTCGCAAAATGTTAGCTTTGCCCGCAAAACGTCATAC 12865
12875 GCCGCTGGAAGAGGCGGCGGTTTCAGGCTTATGTGTGCGGCGCGC 12865
12968 CAACCGATGCCATGCCCGACAGGTGCACAAACTCTTCCGCGCGAGCGGATGCGGCA 13018
13028 GCGCTTTTCAAAATGCTCATGTGATGAACGTGCGGAGATTAATGAAATACGACTTTC 13028
13088 GCGCGGTGCGAAAGTTCATCAGAACCGCACGCGCAGGATTAATGAAG 13077
13229 ATTTCCAAACGGGATTTGCCGATTTGCCCGCCGAGGCTGTTATGATTTGGGATGCGG 13279
13289 CCGAAGCTATTCAGAGAACCTGTGACGATTTTGGCGCCATTCGCTGTGCGGCAAA 13292
13434 CGAAGAAATTAAGAACTGCTGTTTTCAGGGCACGCGCGAGCTGTAAACGTTTGAA 13484
```

```
7978 7983                                8022 8027
8037   TAATCTGTGACGGGGTAGGCAATCAGCCGTTTTCACG
...
8180   TCGATTTGCGTTTCGTGCGTATGTTCAATAAAGGTGTGCGCGCCGCCGCGGTGCCC
      8230 8235
8240   GTGGTCTGCATTCGAGATTTCGCGCGCGCTGTAACTGCTGAGGAATAATGCTTCGACAATG
      8332 8337
8300   CCGTTTGAATCAGCAAGCCGCGTGCCTGAGCAACGCTTCGCGCATCAAAATAGTGTG
      8332 8337
8360   CGGAAAGACGGGGGATGTGCGTTCTT
...
8409   GCAGGACTTTTGGCGATGCTGTGATCAGGAAACTGCTTGGCGGTAGAGGCTCCGC
      8459 8464
8469   CGGAGATGCGCCGACGAGGTGTCCGATAAGACCGCCCGAAACGGT
...
8633   CAGGCGGAATCTACCAAGTAGTCGGCGTGCATGCCGTTTTCGTGCGCGGCAACGACGTG
      8683 8688
      8686 8691
8693   CAGGAATGCTGTGTGCTGCTTCCGCTGCGGCAAAACCGTAGG
...
8846   AAGCGACGGGCGCTTCCGTATCAAAATCCCATTCGTGTTAAAGTCGGGGGTCGCCGATG
      8896 8901
8906   TGTGGCCCATCAACTCGGGGTGCGCAAGTCGCGCGCAACGCTTTCGCGGTGTGCGG
      8965
8966   -----|
      8970
      9068 9073
9026   CCGGCGCGGCTTTTGCGTTTGCCGACGTAACGGTAATGTCGAGCGACTTGTCTCTGA
      9068 9073
9086   AACTCGATTGTTCGATTTCGCCACGCGCACGCTGACGCTTGTGCCAATGATTGCTG
      9121 9126
9146   AAATCGGCTTCGCGCGCGGTTGCGCCGCTGCTTTTGGCCAAGTCGAGCGTGGCGGCGAG
      9173 9178
9206   AGGTGAGAGTTTGGAAACGGT
...
9443   AGCAGATGACCGTTTGAGAGATTGGATGAACTGACCAAGCTCTCAAGCATACGC
      9493 9498
9503   TGAATAAATCGGCTTGACGAGATTTGTACGAAGCCGCTTACCTATAAAAAAATTA
      9541 9546

9563   CGTCAACGGCGGCTCAAGCGGACGCGCAATTATTCGCAAGCTGATGCGGACACGG
      9639 9644
9623   ATCCCGCGCCATCGAGGGCTTCTTGGCAAGCTGCGCGCGACGATGCGCGCACACAG
      9661 9666
      9664 9669
9683   -|
      9684
      9727 9732
9743   CCGAGTTATGTGSAATTTTCGCGCATGCGGACGCGGTAATTTGCTTACCTGATCCGCA
      9772 9777
9803   ACACCAAAAAAGACGAGCAAAA
...
9871   ACCGTTATGAAAGCCAGGCGGACAGGGAAGCCTAGGCATGTTTTCAGAGACATCC
      9921 9926
9931   GCGGTTATTAGATTGAGGATAAATGTTGTTCCGTAAGACGCCGCGGCTTTGGC
      9970 9975
9991   GGCACCTTGATGTTGAACGGCTGTAGGTAATGATGTGGGTATGACAGCCGTTTCAG
      10054
10051   CGAAACGACCGCGCCGCAACACAGTTGACAAGACCAATCCGCGCCTTGCGTGTGTTG
      10054
...
10122   CCCAATTGGAAMAGGCGACGCTGCTGATGATGAGGCGGGAATACGTGTTGTCGTCAATC
      10172
10182   CTGAAGATTTCGGCAAGCTGTACGGGCAATTTTGAAGCCGCGGTTGA
...
10244   GTTGAAGCCCAACCGCGCTTTGCTTACCAAGCCCTGCGGTCAAACTGCAATGCGCGCC
      10294
10304   AGCCAGAAATTTCAATACGGAAGGCTTTCCTGCGCTACGATACGACAGACCTGCCGAC
      10364
10364   ATCGCCAAAGCTGAACAGCTTGAGTTGAAGCGGTGCAACTGACAAATGGACAT
      10364
...
10425   CCGGCTGCGCTTCGCGCAAGGCAATACTACGCCACACCGCAAAAAGTGAAGCGCGATT
      10475
10485   ATCATTTTGAGCAAAAGTGTGCTGCGGATATTTATACAGGTTAC
...
10583   ACCAGTGTATCTGATGCGGTGCGCGGCTGCTGCGCTTCGCTGCGGCGCTTGATT
      10633
10643   GCAGCCAGCAATTCCTCAGACAAATGACGGAATGCCGTCTGAAA
...

```

```
4869 GGTTCGACAAAGGTGCACAAACGGATTAATCGGTCACAGCGGTGGGAACTCGGGCGTG
4929 AAGAGCTTGTGATTTGCTGCATTTGTCGCCCGCTTGCGCTTGCGGTAAGCCGAGG
4989 CTGGGCGGCTGAGGCTTTCCCGCCCGGTTGTGTGTCATGATTAAGATGACGTTGGCGG
5049 AAATCGGACCTTACCGCTTGTGTGCGTCACTTGCCTCCATGACTTGACAGAGAG
5109 ACGTTGAAATGTGGGGTGGGCTTTTTCGATTTCTGTCACAGACACGCAAT
5204 TTCAAGCCGAGTAGCCCGGTGGCGCCGATGATGCTGATACGGCGTGGCTTCAT
5264 GTATTCGACATATCAAAAGCGTTGCGAGCGGTACGCCCATACGAGTAGGCAAGCTGTTTGGC
5324 GACTTCGGTTTTGGCCGAGCGCGTGGCGGCGGAGAGAAGAGAGCTCATGGGTTTGTGTC
5384 GGGCAGCCCA
5451 CCGTAACCATGTTTTTCAAAATCGGCGCGGAGAAATTCAGACACTTCTGTCGTCGTC
5511 GACACGTTTTTTTGGGGAATCCGCGCAGCTTTGGCAATGACGGTTT
5609 TTGGCGCGCGCTGCTTCGTCATTAACATCGATGCTTTTTCGGGCAAGAAAGCTGCTT
5669 GATGTAGCGGGCGAGAGTTGCGCGCGCTTCGATGCGGCTTGC
6000 TCGGGATTTTTTTCGAGCTGTTTCAGACGATTAACCCGCGCTCAAAAGTCGCGCGG
6060 TATTTGTCGCCGCCCAAAAGCGAATCATATCAAGCGGTACACTTGGCATCTTTAAGC
6120 GCGTCGGGAATGTCGCCCTTGACGATTTGATGTGCCAAACCTTCCGCCAGCGCGTTTGG
6180 CCCACACCGCGCTGCG
6252 ATTTCGTGTTGCGACCATCAAAAGGTGATACGCCGCTTGACTTGGCGCTTGAGG
6312 TTGACGCTGTAGTCGAAAGAGACCGGTTTTTTGTTCCGCTTCTT
...
6600 GTTTGACCGTATCCAGAAAGATGATCGGAATTTGGCGCTGTTTTCTGCACAGCGCGG
6660 GCAAGTTGCTCGGACAGACTTTTAATTCGACACACAGATTTTCAAGAGCTTGCGGACA
6720 GCCGATCTTCTTATCAGTACCAAAAGCAGGTGTTCCAGCTGATAACTCATTAACG
6780 GCGTTGCGTCTTACAGATAGAGCAGTTGACAGATGTGTTCCAAATTCGGGTGAAA
6981 ACGCCACTGCTGTCTGTGTCAGCAATAAAGACTGCTCAGGATTTTCAGACCAAT
7041 TTCATGTGTGTATTCGTTCGTTCAATAAATACGCCGTACGTTTGCGCGCGCTGTGTG
7101 TTGATATGCGTACAGAGGTATCGATGTGTGTGCTGATTCGATGCTTCAATATCTTA
7161 G
7210 GTATTCGCAATTTTTTTCTATTTTCCACTTTTGTGAAAACATGCTTACGCTTTTGT
7270 CTTAACAATGTAAAAAGAGGTTAAGCAGAGATTGCACTGACTGCTGATGCATACG
7330 TGGAGAAACGCTGAAAGCTTTTGTAGTTGCTGTATGCTTGTGCTGATTTTGTAAAT
7390 TTTAAGTATAGGAATTTCTAATGGAACCGGTATCGTAAATGATGTTTAAAGAGCTTAA
7450 GGTTTTGTTCATCACTCTGATGAAGGTGCGAAGATTTGTTGCTCACTTCTCAAGC
7510 ATCAACATGATGTGTTCAAAACCTTGAAAGAAGTCAACCGTATCTTTCG
7577 GCAGGACAGGGGCAAAAGTGGCCCTTCTGCTTCAGACGCAATGCGCAGGCGCTTCAAGC
7917 TTCCGGAACCGTCATCCCGCAATCAGAAATCGAGCGGATTTGTGACAGCAAGCCGCA
7977 AAGCGTATCCGCGACCGCCAGCAATGTGCGGTTACATATCTCGAAGCTCCGCGTACGG
```

```
1815 TCAGATTATTCATCAAGTTTGGACGACGGGACGATGCAAAACCGCTATGACGTTACCG |-----|
1875 TGTGTGTAAGGCGCTGCCGCTGCTGAGATTGAATTGAATAAGCG |-----|
...
2303 AGCAAAACCGGAAAGCGGCGGTATGTTGGCATACACGGGCGACGGCAAAAGCGCTGACC |-----|
2363 AGCTTTAAGCGGGCGCTCTGGCGACGGAATCGCGTTTATCGACAAGGTTTCTCTGCTG |-----|
2423 GTGGACAGAAAGATTGGAAGTATCAGACGATGAAGAAATACCAAGCTTTTTCGCCGAC |-----|
2483 AGCGTGAATGTTTCGAAAGCAGCGGCGGCTTGAAGCAATTTGAAAAAGAGACAAC |-----|
2543 AAAATCATCGTTACACACCATCCAAAAGCTGAACAACCTGATGAAGAGTGAAGATATCTG |-----|
2603 CC |-----|
...
2690 TTAAGAAATTTCTGCAATTTGCGTTTACCGGTACCGGATTTTCCGAAACGCTTTGG |-----|
2750 GCGCGAAACACGCGCGGCGGTGTTGCGGCGTGAATTGCATTTCTTA |-----|
...
2801 TTACCGATGCCATCCGCGATGAAAAAGTATTGAATTCAAAGTGATTAACAGATGTGC |-----|
2861 GCCCGAGTTCAAAGCGGTGGAAGCGGAGACAGATGAGAAAGAACT |-----|
...
2978 ACCAGTTCAAGCAGAAAAACGACCGCTGATGCGGCGGTAAAGGCTTCAACGCCATGT |-----|
3038 TTGCGGTCAAGCAGCGTGAGATGCGCGAAGTCTATTAGAAAGCGTTCAAAAACACACAGG |-----|
3098 CAGCAGCCTGCACCGCGCTGAAGTGGCCACCAATTTTTCCTTGGCGGCAACAGAGAGC |-----|
3158 AGAACGCCGCTGGTGAATTTGTGATGAGACCTTTGAACCGGAGCGATGAGACGACGCG |-----|
3218 CGAAAGATTTTTCGAGGCTGCCATCAACGATTAACAAGCGCTGTTCAAAACCAATTCG |-----|
3278 GCACGACAGCAAAAGCCTTTCAGAACTACTTA |-----|
...
3343 GTGGATTTGCTGATTTGTGTGGTATGTTTTCAGCGGTTTTCAGCGCGCGAGCGCTGAAC |-----|
3393 3398
```

```
3403 ACGCTGTTTGTGATATAAAACCTGGCGTATCAGCGCTGATGCAAGCGCTTTTCGGCACC |-----|
3463 AACCGCATTTTACATGCCACCAAAACCTTGGCAATATTG |-----|
3503 TCTGCTCCGGAATTTGGAGCAGCAACCATTTGATGCGATTACCTGTTTCGGCATATAA |-----|
3563 ACACCAAAAACGTGTGTTGAAAAAAGTACGAAATACATGA |-----|
...
3627 GACGGGCGAAGTGGCGCGGCTTATTGGATGTGCAAAAGATTTCCGCGAGCTTCCC |-----|
3687 CGATCCCGACAAAATCGAAACGGAAGCAAAAAGATTTTGCC |-----|
...
3793 GCGCTGGCGAGTTGCAGAGTGTGACGCGGCGGACGAAAGATGCGATGAAGCGCTTCAA |-----|
3853 GAAAAATACTACTGAGCGATGAAGACGTGCAGAAATCGGCAAGTGCATGCCGCTCT |-----|
3913 GAAAGGCGAGTGCAGACTACCGTTTCCCTATACGACATCCGCGACTGGCTGGCGCCG |-----|
3973 CAAAAACGAGCGAAGCAAGAAAGAACAAATCAAAAAATCGATTTGGACAGTGTGTTTTGAG |-----|
4033 GTGATTTGCTCAAAATCGCAGAAATCATCTGATTA |-----|
...
4151 GCGCGACGATCGCGCACCGCGCCAAAGAGACCTGATTTGTGATTTTCATCAGCATACGG |-----|
4211 ATTTGACAAAAGTCCCGAGCTTCCTGCAATCTGAAACCTTTTACACCTACGCTCAAG |-----|
4271 AGTGATGAGGCG |-----|
4284 CGAAGCAGCAGAGCTGATTGCCCGCAAGACTGTAACGAAACCGCCGCCAAACGTTAATT |-----|
4344 GACCGGCTCGCTCAAAAGCGGCTATGCCAGGAAAGCGCAGGAA |-----|
...
4420 CTCAACCCGCAATATCTGACGAAAGAAAGTGTTTTTCAAAGATTTGGCGGTTTTG |-----|
4480 GAAAAATTTGCCGGAATAGGACAGAAATTTAAAAAATGCCGTCT |-----|
...
4809 TTGTCGAGAGACCGGTGTTGAGCTGAGAGAGAAATTTGTCACGACTTTGACGATATATG |-----|
4859 4864
```

```
CDS
/gene="NMA1436"
/327992..328771
/note="NMA1436"
/len=259 aa; unknown, rich in Ser, Thr and
Asn, lies within a region of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1436"
/protein_id="CAB84672.1"
/db_xref="GI:7380088"
/db_xref="SPTREMBL:Q9J094"
/translation="MRYVNTISPPDENKGSISRSKTKGNLFNNVNVVHKOTHHGCH
NLNNTDVHLAEDSDNVVAIKGELADPPTGNVESTFLKGIETAFKSNVTGNNK
ISIAEGKGLGVNGLOSNSPSGHTTGIFINQSSDLPGLDSTIKLKAVKGN
NEVDLNTVNNHASVKGIIASHSAKATLEAGDNITVKNKDPNTALIDALKQYPAK
ELRYGVOSTGTGVYISPNKIALTELTFGIVGICVSL"
complement(328767..328921)
/note="Correia element; hmms hit to HMM Correia (1 -
156), score: 276.86"
/label=Correia
complement(329043..329957)
/gene="NMA1437"
complement(329043..329957)
/gene="NMA1437"
/note="NMA1437, probable lysR-family transcriptional
regulator, len: 304 aa; shows weak similarity to many e.g.
SW:XAPR_ECOLI (EMBL:X63976), xapR, Escherichia coli
xanthosine operon regulatory protein (294 aa), fasta
scores: E(): 1.6e-11, 23.7% identity in 295 aa overlap and
TR:030441 (EMBL:AF006000), brgI, Bordetella pertussis
transcriptional regulator homologue regulated by the
oxygen-responsive transcriptional regulator btr (291 aa),
fasta scores: E(): 5.4e-13, 28.7% identity in 296 aa
overlap. Contains probable helix-turn-helix motif at aa
17-38 (Score 1810, +5.35 SD). Contains Pfam match to entry
PF00126 HTH_1, Bacterial regulatory helix-turn-helix
protein, lysR family and PS00044 Bacterial regulatory
proteins, lysR family signature"
/codon_start=1
/transl_table=11
/product="putative lysR-family transcriptional regulator"
/protein_id="CAB84673.1"
/db_xref="GI:7380089"
/db_xref="SPTREMBL:Q9JUR3"
/translation="MDAVQLKSFVAVHAGNLTQAAKRLFLSPAVSAQIKALEEYVG
TPLFRRTGKGVLTFRAGEILPEASELSQYHKLHFAKTLAGDVEETSLGIIHPID
SAKIVALTNDNGOFTPKTRLIHQVGMGEILSRIOHKTLLHGFTLGNAAORGIREVL
ONTLYALICPOSOYPHLTRSLPSLOEGCWTEMSGVSGSRKHILQFMRSNRLSPKQI
LCDYPTQIIDIYAGSIGVAMPKNAKAAKAGAVAIIESCRHSMPLNFTYAEIED
NPVSLLECTIEKWGVQAVQPPVVDN"
complement(329523..329951)
/gene="NMA1437"
/note="Pfam match to entry PF00126 HTH_1, Bacterial
regulatory helix-turn-helix protein, lysR family, score
117.10, E-value 3.4e-31"
complement(329617..329909)
/gene="NMA1437"
/note="PS00044 Bacterial regulatory proteins, lysR family
signature"
complement(329963..329968)
RBS
330201..330206
RBS
330213..331682
gene
330213..331682
/gene="NMA1438"
330213..331682
/gene="NMA1438"
/note="NMA1438, possible integral membrane protein, len:
489 aa; unknown, contains hydrophobic, possible
membrane-spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB84674.1"
/db_xref="GI:7380090"

misc_feature
/db_xref="SPTREMBL:Q9J093"
/translation="MEKHNGTYRDLIHRPASEFATRDEXLEHLOIMQPKRMRNLPFR
DYREMEDILIPAMAGTICKVYVWVGAFAFAPAGLPSPFLVENVRYELLIAAFTIL
VSGFELPGANLPCTHGPILIPMIPIVYVSAAGHPLAFGISIVLGLMALPFGSGTAKL
TSNGYCGSLILYGFITGTTGQVKKLFSSNAGFPNNPIAFYIYITVYALLENHKKR
WLAVPLGSLINGVAALFALGAPPEHETKISYPERKTVMSPAVWGENSGWHLGLPTAESFLV
FEPFALVAVMNSPDELGHQVOKISYPERKTVMSPAVWGENSGWHLGLPTAESFLV
FTSSWGTIIVPASIAKRPIDPGAVLTAVLCIIAGLIMQVMDLAIWQPYVALVGVY
LPLEAGEMETRKGTQSAATVSSALVNPVFGMALTYMLLDNLGLGCKERSAQLG
FTGRVLIIPVGEFLICVMAVGMIPGIPPLEHFKSLG"
complement(331713..331722)
/misc_feature
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
BASE COUNT
79079 a 81425 c 89192 g 82105 t
ORIGIN
NMA422491 Length: 331801 February 21, 2002 16:59 Type: N Check: 8990 ..
Found using 'seq2-3' (pappu403.key)

596
AAAAATACCTTGACAAATTCACCGAGCTGGAACCTGAATTAGCCCTGCCGAACGCCATA
-----
656
CCGATATTACCGCGACTTCTTTTAGATTGACAATCAATCGGGGGGGGATAGCTGAT
-----
716
GGCTATAAGGCGCTCTGAAGATGTGTTGGAAGACCTTGGGGAGGTGTTGACTTG
-----
776
AAGAAATGCTATACCCCTTCCAAATCAATA
-----
1218
ACAAATTCGATACCTGACCCACTCTATACGGAAGGCGCTGCCGACGAATGCCCTGTC
-----
1278
GCCGCAACATATACGATATTACCGGAAACAGCTGCTGCTTCCC
-----
1357
CAGGACGCCCAATCCGAACATCAAAACCGTTTAAACCCGATCTGCAAGGAGCAGCCAT
-----
1417
GAACCTCGAACAACCAACCCATGCGCGAAGGCGGATTTTCATGCTGACCAATATGGA
-----
1477
AAAAATCGAACAGTGGGAGCTACCAATCGGAACCAACGATTGGAAGCGGAGATTATGCG
-----
1537
CGATTTGCAGAATCAGGATTACGAATACCGCAAGATTGGAACAGCAACAGCGCTGCT
-----
1597
GGAANAACCTGCGCGCGAGCTGACGCGGCTGACGATGTGGCTTTTCAGACGCGGAAATG
-----
1657
GGCGGCGTTTTCAGCAATATCTGGACAGCGCGTCTGAAACATTTACGATAAAAACCGG
-----
1717
CAAAATTCATGACGACCATATTATGATTGCTTTGATGACGCGCTGAAAAACAT
-----
1777
TTATCT
-----
```

transl_table-11
/product="putative bifunctional 3,4-dihydroxy-2-butanone
4-phosphate synthase/GTP cyclohydrolase II"
/protein_id="CAB84667.1"
/db_xref="GI:7380085"
/gb_xref="SPTREMBL:09J097"
/translation="MSHSIRPEIILADIRKVMYIITAEEDRENGDLMAAGVTPPE
AINEFKHARGVLCIPMOENVEKELGLPMPTOTKNAQGTNETVISEAHITTTGIS
ABRALITIAVSPTAKPEDIVQPGHIFPLRAKGSVIVRAHTEAGVDLAOMNLIPRA
AICETILINDGSMAMPPELMKFAEHHKLGITIDILIEYRSRTESILEDMAONAVQTRP
MEEOOHVYVDLISGETHIALKVGPSADTEILVHVEPESVMDIQANPHSMDLPE
ALFEVQVAGASGVILHHRDEGASILDPTPLPGANQATKMKSKSGICAGIOLALANK
KIRVIGQSPSPFGITGFGLEVGFEEAEK"
32737. .32324
/gene="ribB"
/note="Pfam match to entry PF00926 DHP synthase,
3,4-dihydroxy-2-butanone 4-phosphate synthase, score
415.20, E-value 6e-121"
32331. .32398
/gene="ribB"
/note="Pfam match to entry PF00925 GTP_cyclohydrol2, GTP
cyclohydrolase II, score 12.90, E-value 1.1e-08"
32393. .323950
/note=">= 90% match to ARTCCNNNNNNNGCAAT"
/label=dRS3
323951. .324006
/note="82% identity to consensus
CCAGCCGCTCAGTTTCAGTCATTTCCGATTAATTCCTGTTTCATTTCTAG"
/label=RS17
324035. .325253
/note="IS1106A3 remnant"
/label=IS1106A3
324035. .324053
/note="Inverted repeat at end of IS1106A3"
/label=IR
complement(324187. .325190)
/gene="NMA1430"
complement(324187. .325190)
/gene="NMA1430"
/note="NMA1430, pseudogene, probable transposase, len:
1004 bp: highly similar to many e.g. TR:CAB44967
(EMBL:AJ242841) Neisseria meningitidis strain 22491
putative transposase for IS1106A3 (335 aa), fasta scores:
E1): 0, 97.8% identity in 224 aa overlap (frame 6) and
E1): 0, 96.3% identity in 109 aa overlap (frame 1). The
potential coding region frameshifts at codon 109 approx.
near a cpg(12) repeat"
/codon_start=1
/pseudo
/transl_table=11
/product="putative transposase (pseudogene)"
complement(325253. .325253)
/note="Inverted repeat at end of IS1106A3"
/label=IR
325279. .325313
/note="Stem loop containing DNA uptake sequences: aaaa
ggcgcctgaa agcgt ttacagacgc atttt"
325284. .325293
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
complement(325348. .325662)
/gene="NMA1432"
complement(325348. .325662)
/gene="NMA1432"
/note="NMA1432, pseudogene, possible type I restriction
enzyme system modification protein, len: 315 bp: similar
to parts of putative type I restriction enzyme system
modification components e.g. TR:G92M08 (EMBL:AE001475),
HSDM1, Helicobacter pylori j99 type I restriction enzyme
modification subunit) (543 aa), fasta scores: E1):

```

misc_feature
1.1e-07, 38.7% identity in 106 aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="putative type I restriction enzyme system
modification protein (pseudogene)".
325681..325835
/note="Correia element: hmms hit to HMM Correia (1 -
156), score: 284.48"
/label=Correia
complement(325932..325941)
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
complement(325948..327258)
/gene="NMA1433"
complement(325948..327258)
/gene="NMA1433"
/note="NMA1433, len: 436 aa: unknown, similar to many
hypothetical proteins e.g. SW:YCAJ_HABIN (EMBL:U32833),
H11590, Haemophilus influenzae hypothetical protein (446
aa), fasta scores: E(): 0, 54.0% identity in 437 aa
overlap. may have some role in DNA replication or repair
as the N-terminal half is similar to yeast replication
factors e.g. SW:RRC3_YEAST (EMBL:U18755), rfc3,
Saccharomyces cerevisiae replication factor C 40 kD
subunit (340 aa), fasta scores: E(): 1.4e-10, 30.5%
identity in 233 aa overlap and to bacterial helicase
components SW:RUVB_THEMA (EMBL:U38840), ruvb, Thermotoga
maritima (334 aa), fasta scores: E(): 1.1e-07, 28.3%
identity in 244 aa overlap. Contains pfam match to entry
PF00004 AAA, ATPases associated with various cellular
activities (AAA) and PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1433"
/protein_id="CAB84670.1"
/db_xref="GI:7380086"
/db_xref="SPTREMBL:O9JU96"
/translation="MTDLFARFAPYAPLAERLRPHRLDDPVGOEHLIGEGKPLRAVAREGE
GPRHSLMLGPRGPGVKTLRLAQSNAOPLPAPSVASGVKDJREAIKRAELAIOOG
RATILFVDEVHRFNKAOODAPLPHYESLLPIFGATTENPSEFVNNPALISAQYVYVIO
PLISDDKLKLAKVALPEYDEFTTETDAQULVNTVDGSRRLNLLEQLRABRDRG
RLKNTLAERLADLSGAQIRFEDKGESEYNOISALHKVSRRSHNALWYFCNRADG
TDPRLAIRIVRIAMEDIGLADPRALQIANDAAATFERLGSPESELALQAVLYIAAA
ASNGYAVRANOMRHFVKNASDEVPVHLRNAPTKIMELGEGREYRVAHDEPNAAYVA
GESWPDGIDEPDEYOPAPRGIEIKIGKRLAMKSIDDEVYAKAR"
complement(326596..327114)
/gene="NMA1433"
/note="Pfam match to entry PF00004 AAA, ATPases associated
with various cellular activities (AAA), score 19.20,
E-value 8.8e-07"
complement(327076..327099)
/gene="NMA1433"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(327264..327267)
327507..327992
/gene="NMA1435"
327507..327992
/gene="NMA1435"
/note="NMA1435, len: 161 aa: unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1435"
/protein_id="CAB84671.1"
/db_xref="GI:7380087"
/db_xref="SPTREMBL:O9JU95"
/translation="MLIOSALCFEAEAEYTNIOEMEWLSFSPOKDDSLITLMKE
KITVYAIYPMKKKGEMFRKRLILITGGKNGKNTFVNGICYGISFAPRODGGYPRLSV
KSNNDILFRNNIQSVDYKPHSGNTHLDKKGKTIQPNLNVSDGDGVDAASGKTRITTV
P"
327992..328771

```

```

misc_feature
320491..321000
/gene="r1ba"
/feature="Pfam match to entry PF00925 GTP_cyclohydrot2, GTP
cyclohydrolase II', score 347.50, E-value 1.5e-100"
complement(320607..320616)
/misc_feature
/note="Core DNA uptake sequence: gccgctctgaa"
/label=DUS
gene
321086..321205
/gene="NMA1426"
321066..321205
/gene="NMA1426"
/note="NMA1426, len: 39 aa; unknown"
/codon_start=1
/misc_feature
/transl_table=11
/product="hypothetical protein NMA1426"
/protein_id="CAB84665.1"
/db_xref="GI:7380084"
/db_xref="SP|REMBL:Q9JU98"
/translacion="MCGTCILRLRLSGNRPFLDKAISPFLQGDANALRRV"
321245..322255
/gene="NMA1427"
321246..322255
/gene="NMA1427"
/misc_feature
/gene="NMA1427"
/note="NMA1427, probable glycosyl transferase, pseudogene,
len: 1010 bp; N-terminus similar to the N-termini of many
e.g. TR:007340 (EMBL:X85787), cps14J, Streptococcus
pneumoniae SS-1,4-galactosyltransferase (318 aa), fasta
scores; E(): 6.2e-17, 52.7% identity in 91 aa overlap.
C-terminus shows weak similarity to many e.g. TR:087183
(EMBL:AB017355), cpsI, Streptococcus agalactiae
galactosyltransferase (315 aa), fasta scores; E(): 0.024,
24.6% identity in 191 aa overlap and TR:007340
(EMBL:X85787), cps14J, Streptococcus pneumoniae
SS-1,4-galactosyltransferase (318 aa), fasta scores; E():
1.3, 25.4% identity in 138 aa overlap. Contains Pfam match
to entry PF00535 Glycos_transf_2, Glycosyl transferases.
Contains a q[9] repeat which would allow translation as an
infect CDS, if variable"
/codon_start=1
/misc_feature
/pseudo
/transl_table=11
/product="pseudogene (putative glycosyl transferase)"
321255..321602
/gene="NMA1427"
/misc_feature
/note="Pfam match to entry PF00535 Glycos_transf_2,
glycosyl transferases, score 39.90, E-value 5.8e-08"
321551..321559
/note="(9)9"
322324..322367
/note="Stem loop containing DNA uptake sequences: gcaat
gccgcctctgaa gccacgaatccgc ttcacagcgc atctg"
322329..322338
/misc_feature
/note="Core DNA uptake sequence: gccgcctctgaa"
/label=DUS
misc_feature
complement(322353..322362)
/note="Core DNA uptake sequence: gccgcctctgaa"
/label=DUS
RBS
322706..322709
322719..322810
/gene="r1bb"
322719..322810
/gene="r1bb"
CDS
322719..322810
/note="r1bb"
/note="NMA1429, ribB, probable bifunctional
3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP
cyclohydrolase II', len: 445 aa; similar to many e.g.
SW:RIIB, PHOPO (EMBL:L11391), ribB, Photobacterium
phosphoreum 3,4-dihydroxy-2-butanone 4-phosphate synthase
(363 aa), fasta scores; E(): 0.47, 8% identity in 360 aa
overlap. Contains Pfam matches to entry PF00926
DHBP_synthase, 3,4-dihydroxy-2-butanone 4-phosphate
synthase and Pfam match to entry PF00925 GTP_cyclohydrot2,
GTP cyclohydrolase II'"
/codon_start=1

```



```
misc_feature      E-value 1.1e-05
310329. .310352
/gene="ruvb"
/stem_loop        311152. .311193
/note="Stem loop containing DNA uptake sequences: gcaat
gccgtctgaa acagagctaatt ttcagcgcg atttc"
311157. .311166
/note="Core DNA uptake sequence: gccgtctgaa"
/misc_feature     /label=DUS
gene              complement(311173. .311901)
/gene="rpe"
complement(311173. .311901)
/gene="rpe"
CDS              /gene="rpe"
/EC-number="5.1.3.1"
/note="NMA1413, rpe, probable ribulose-phosphate
3-epimerase, len: 242 aa; similar to many e.g.
SW:RPE_ECOLI (EMBL:219601), rpe, Escherichia coli
ribulose-phosphate 3-epimerase (EC 5.1.3.1) (225 aa),
fasta scores: E(): 0, 66.2% identity in 225 aa overlap.
Ribulose phosphate 3 epimerase family, PS01085
Ribulose-phosphate 3-epimerase family signature 1 and
PS01086 Ribulose-phosphate 3-epimerase family signature 2"
/codon_start=1
/transl_table=11
/product="putative ribulose-phosphate 3-epimerase"
/misc_feature     /protein_id="CAB84653.1"
/db_xref="GI:7380072"
/transl_table=11
/db_xref="SPTREMBL:Q9JUA9"
/translation="MTYRIAPSLISADPARGVESVIAAGADLHPDVMNHYP
NLTFGPVCAALPYASVPIDVHLMVBPVDLLOSPKAKASITTPPEKSRHDSL
SLTRDWCAGALPYANPATPYLLLENVDRIADNVLMSVNPFGGQSTPYLEKTRV
RAMLDREKSGRIALEVDGIGIKTDNIAAQAAGADTFVAGSAIFGKPKYKAVIDAM
RAELERSGRLLNAPMEIEMPSN"
misc_feature      complement(311179. .311188)
/gene="rpe"
/note="Core DNA uptake sequence: gccgtctgaa"
/misc_feature     /label=DUS
misc_feature      complement(311275. .311889)
/gene="rpe"
/note="Pfam match to entry PF00834 Ribul_P_3_epim,
Ribulose-phosphate 3 epimerase family, score 454.40,
E-value 9.8e-133"
misc_feature      complement(311431. .311499)
/gene="rpe"
/note="PS01086 Ribulose-phosphate 3-epimerase family
signature 2"
misc_feature      complement(311761. .311805)
/gene="rpe"
/note="PS01085 Ribulose-phosphate 3-epimerase family
signature 1"
RBS              complement(311909. .311912)
311925. .312079
/note="Correia element; hmufs hlt to HMW Correia (1 -
156), score: 281.94"
RBS              /label=Correia
repeat_unit      312199. .312204
gene             312212. .312646
CDS              /gene="NMA1414"
312212. .312646
/gene="NMA1414"
/note="NMA1414, possible periplasmic protein, len: 144 aa;
shows weak similarity to SW:SPY_ECOLI (EMBL:Y07714), spy,
Escherichia coli spheroplast protein (161 aa), fasta
scores: E(): 0.0013, 27.6% identity in 134 aa overlap.
Contains a possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/misc_feature     /protein_id="CAB84654.1"
/db_xref="GI:7380073"
/db_xref="SPTREMBL:Q9JUA8"

/translation="WPLPAPCRFAKPAASFLSMALLSCQLSHAATVAVIPLNDPQNC
IRRLGLTGGHNELRKRIIAAFKMAAGDRLKVMHSEHSRRSVVEIISSDVFNREAR
DYVESRYTSGMDFAVDLEIQRFFHITTPQOOOMWLSSICK"
gene             complement(312659. .313522)
CDS              /gene="NMA1415"
complement(312659. .313522)
/note="NMA1415, len: 287 aa; unknown, similar to the
C-terminal half of TR:Q55502 (EMBL:D64006) Synchocystis
sp. (strain PCC 6803) hypothetical protein (436 aa), fasta
scores: E(): 0, 46.9% identity in 241 aa overlap and also
to TR:Q66668 (EMBL:AE000684), AQ336, Aquifex aeolicus
hypothetical protein (196 aa), fasta scores: E(): 3.4e-19,
46.5% identity in 142 aa overlap. Contains PS00017
ATP/GRP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1415"
/misc_feature     /protein_id="CAB84653.1"
/db_xref="GI:7380074"
/db_xref="SPTREMBL:Q9JUA7"
/translation="MEINFEIDKAVAVLRRLDAVLPEPGHTMNALAFPMOSAGKC
FLEHLPPHPPEPLARLAVGVRORFLLVNRDEQIAGRPANNVLSMGKRGKSSVVA
LHLEYAKGRLILEVDKSDILGPLYLTLLKERPEKTFVCDLSEFSGETYKALPT
ALDGLSQRCANVLVATSNRRHLMPYFEDENAGTTGMGEIHQKEAVERKISLDRF
GLMLSFYFDQNDYLAIVONWLEDFGVDETAQMAALQWATGRSGRSASMOFAD
WAGRLPKORAL"
misc_feature      complement(313232. .313255)
/gene="NMA1415"
/note="PS00017 ATP/GRP-binding site motif A (P-loop)"
RBS              complement(313528. .313531)
gene             complement(313530. .314192)
CDS              /gene="ribC"
complement(313530. .314192)
/gene="ribC"
/EC-number="2.5.1.9"
/note="NMA1416, ribC, probable riboflavin synthase alpha
subunit, len: 220 aa; similar to many e.g. SW:RISA_ECOLI
(EMBL:X69109), risa, Escherichia coli riboflavin synthase
alpha chain (EC 2.5.1.9) (213 aa), fasta scores: E(): 0,
62.7% identity in 204 aa overlap. Contains two Pfam
matches to entry PF00677 Lum_binding, Lumazine binding
domain"
/codon_start=1
/transl_table=11
/product="putative riboflavin synthase alpha subunit"
/misc_feature     /protein_id="CAB84656.1"
/db_xref="GI:7380075"
/db_xref="SPTREMBL:Q9JUA6"
/translation="MLQFAFLPTQGNRIWFTIGVIGLRTAIIHRSFAFYVVEL
PQEAADNLQAGASVANNNGCGLTTEIEGNVSFSDLAETLAKNLGLGQCVNIER
AAREGDEIGSHVMSGHIMATPIVLEIERGQVNTVFSIDHEIKPYLILKPGYLDGC
SLTIGKVEDSRVNHILPETLETLFGSKRVGRINIEIDPNTQIVDYVERMAORY
AK"
misc_feature      complement(313593. .313847)
/gene="ribC"
/note="Pfam match to entry PF00677 Lum_binding, Lumazine
binding domain, score 115.80, E-value 8.2e-31"
misc_feature      complement(313884. .314141)
/gene="ribC"
/note="Pfam match to entry PF00677 Lum_binding, Lumazine
binding domain, score 106.10, E-value 6.8e-28"
314208. .314217
/note="Core DNA uptake sequence: gccgtctgaa"
/misc_feature     /label=DUS
complement(314310. .314888)
/gene="NMA1417"
/misc_feature     complement(314310. .314888)
/gene="NMA1417"
/note="NMA1417, possible molybdopterin-guanine
dinucleotide biosynthesis protein A, len: 192 aa; similar
to bacterial mobA genes e.g. SW:MOBA_ECOLI (EMBL:L19201),
mobA, Escherichia coli molybdopterin-guanine dinucleotide
```

```

/gene="NMA1409"
/note="NMA1409, probable ABC-transporter ATP-binding
protein, len: 542 aa; similar to many e.g. SW:TLRC-STHFR
(EMBL:M57437), tlrc, streptomycetes fradise tylosin
resistance ATP-binding protein (548 aa), blastp scores:
Expect 1.4e-45 and SW:YB17-ECOLI (EMBL:AE000184), yb17,
Escherichia coli hypothetical ABC-transporter ATP-binding
protein (530 aa), fasta scores: E(): 0, 65, 6% identity in
543 aa overlap. Contains two Pfam matches to entry PF00005
ABC_tran, ABC transporter and two PS00017 ATP/GTP-binding
site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative ABC-transporter ATP-binding protein"
/protein_id="CAB84649.1"
/db_xref="GI:7380068"
/db_xref="SP:REMBL:09JUB3"
/translation="MISTNGITWQFGAKPLFENVSVKFGEGNRYGLIGANGSKSTFM
KILGDLQETAGVAIEVNGVRLGRKODPAYEDMRVLDDVMGHTEMNAATERDI
VANPEATEDDYKAAELKFAEYDGTAEARAALLSGVISEDHNAATMAEVAPE
KLRLVLAOLAFSKPVDLLIDEPNNIDINTIRMLFGVINYDSTMIISHDRIANPV
CTHMADIDYNTITIVPRNYDDVYMLASQSRERAKKNKAKKEKIQEIEVFARFSANK
SKARQATSRUKQADKIKSEMEVYKPTSRQNYTRFEADEKAKLHQAIVEKILAKRE
TQLEKLNFTLEAGORLAIIGPAGAKSTLLKLAGAYNPEYSDGLLPDGSIKMAEK
ASVGYYPQDHENDFDVMDLSEMMRWQGDGDEQVIRGTIGRLFLGSDNVKRVKL
SGEGRMLYKGLLLKPNVLMVDEPTNHMDIESLNMALKEVNGTLIFVSHDRQV
VSSIAQTIELDQKGGCYHYLGDYSEYLEKKVA"
/306631..307235
/gene="NMA1409"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 127.70, E-value 2.1e-34"
/306652..306675
/gene="NMA1409"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/307403..307412
/gene="NMA1409"
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
/307594..308118
/gene="NMA1409"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 149.50, E-value 5.7e-41"
/307615..307638
/gene="NMA1409"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/308194..308228
/note="Stem loop containing DNA uptake sequences: acaat
gccgcttgaa gccgc ttccagcgc attgl"
/308199..308208
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
complement(308214..308223)
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
/308244..308249
/308253..309506
/gene="cca"
/308253..309506
/gene="cca"
/EC_number="2.7.7.25"
/note="NMA1410, cca, probable trna nucleotidyltransferase,
len: 417 aa; similar to many e.g. SW:CCA-ECOLI
(EMBL:M12788), cca, Escherichia coli trna
nucleotidyltransferase (EC 2.7.7.25) (412 aa), fasta
scores: E(): 0, 54.2% identity in 404 aa overlap"
/codon_start=1
/transl_table=11
/product="putative trna nucleotidyltransferase"
/protein_id="CAB84650.1"
/db_xref="GI:7380069"
/db_xref="SP:REMBL:09JUB2"
/translation="MQYLVGAVRDYLLGLPVKKRDWVVGADQTMLAGFQPYGK
DEPVLHPETHREVALARTERTAKGVGFSFHADKDYTLLEODLMRDLITNAMQDA

```

```

DGIIDPFGGORDIAGILRHVSPAFADPYRLITRTAFARYFEIAEETIKLMROM
VENEADALVHERWQEFKALMEKRNPKMIEVREGCALVLLPEVALGVPORAD
YHPEIDSGIHTMLTQRAADMGSLPXYRALLIDIGAKRPSDIPHHHGDVAGVE
PVKRVNRLRARKFCAELAEIVCRMHITTFHOYOLKSOTITLVAKTKDFAFRPERFOT
ALANCIADTQGLRREHETPPYQRAHWALLLEAANOADSGKIAACRSOGKHAHLAEQI
DRARLADJAPLOKAFRAODTEKH"
complement(309247..309256)
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
/309286..309295
/gene="cca"
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
/309528..309589
/note="Correia element; hmms hit to HMM Correia (1 - 62),
score: 99.50"
/label=Correia
/309591..309632
/note="Correia element; hmms hit to HMM Correia (114 -
156), score: 56.34"
/label=Correia
/309729..310037
/gene="NMA1411"
/309729..310037
/gene="NMA1411"
/note="NMA1411, len: 102 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1411"
/protein_id="CAB84651.1"
/db_xref="GI:7380070"
/db_xref="SP:REMBL:09JUB1"
/translation="MTIKTDLPRINNEEDYQRLIKSAEFSGGEIPLNLEKFN
DVVQADLAQAVMQQVRFDPVAIHDSDDDEDTGICPHCINPMPPLRDYLVWRETRG
"
/310060..310102
/note="Stem loop containing DNA uptake sequences: tcaat
gccgcttgaa acgcgcgcgcgcg ttccagcgcg atacc"
/310065..310074
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
complement(310088..310097)
/note="Core DNA uptake sequence: gccgcttgaa"
/label=DUS
/310119..311150
/gene="ruvB"
/310119..311150
/gene="ruvB"
/note="NMA1412, ruvB, probable Holliday junction DNA
helicase, len: 343 aa; similar to many e.g. SW:RUVB-ECOLI
(EMBL:X07091), ruvB, Escherichia coli Holliday junction
DNA helicase (336 aa), fasta scores: E(): 0, 71.8%
identity in 333 aa overlap. Contains Pfam match to entry
PF00004 AAA, ATPases associated with various cellular
activities (AAA) and PS00017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=11
/product="putative Holliday junction DNA helicase"
/protein_id="CAB84652.1"
/db_xref="GI:7380071"
/db_xref="SP:REMBL:09JUB0"
/translation="MTQDNTLTAAQOPRIYAQATASQOELELRLAPKTLIDYIGCH
KAKEQIAIFIOAKKKRREALDHYLFGPGLGKTTLHAIKELGVNLROTSGPVLER
AGLALALTNLDHVDVLEIDEIRLSVVEILYPALEDYRLDIMIGEGPAASVAKID
LPPTLVGATFRAGMTNPLRDRFGIYSREFYENRDLATVRSQQLQDLSSEGA
EETAKSRGPPRIANRLRLRRVDFADKNNGTIDGTAALASMLDVOGLVDMRK
ELEAVLHKGGGPGVGDVNAALIGESDITLEDVIEPILIQGFLQRTPRGRMATERAY
LHRGLPYEK"
/310314..310853
/gene="ruvB"
/note="Pfam match to entry PF00004 AAA, ATPases associated
with various cellular activities (AAA), score 5.00,

```

```

/gene="NMA1402"
/Label=DUS
/misc_feature complement(301827..301850)
/gene="NMA1402"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/complement(301966..303213)
/gene="NMA1403"
/complement(301966..303213)
/gene="NMA1403"
/note="NMA1403, possible integral membrane protein, len:
415 aa; unknown, similar to many bacterial hypothetical
proteins e.g. SW:YCFW_ECOLI (EMBL:AE000212), yxjW,
Escherichia coli hypothetical protein (414 aa), fasta
scores: E(): 0, 35.3% identity in 405 aa overlap. Contains
hydrophobic, possible membrane-spanning regions"
/codon_start=1
/transl_table=1
/product="putative integral membrane protein"
/protein_id="CAB84643.1"
/db_xref="GI:7380062"
/db_xref="SWISS-PROT:P57061"
/translation="MFSLEAMIGLRYLRKRGFMSEFTMVSIGALGVTAIVL
SVNMFQKEIRGOLNVAHPAEIGYIDNTDDMRNLFTENRKGIILAAAPYSNOAL
LANGGEIVQIRGILPSEERKIVVEYDMPAGKRPEDLIRPEFDIILVGLAEALGAE
VGNKYVTTPPGNTPRAGVPRKQFTYVGLKGVTVEDNSLATHIQDARLVRLD
KEVAGLRKLADPQAPALITAKLIPDAQRDVWMDWFSNRSYEAVALRKRMFTII
LTLLIAVAFLNLSVLVAVTEKQADIALIRTLGLSPGVAKIEMVQAGFSGFETLA
GVCGVLLGMVNGRVAVFENLIGVHLINSQVYFIDYLPDVMGDVALINCISIGLS
FVATLYPSRRASKTOPAEALRE"
/complement(301983..301987)
/gene="NMA1403"
/complement(302851..302860)
/gene="NMA1403"
/note="Core DNA uptake sequence: gccgtctgaa"
/Label=DUS
/gene="NMA1404"
/303285..303410
/303285..303410
/gene="NMA1404"
/note="NMA1404, len: 41 aa; unknown"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1404"
/protein_id="CAB84644.1"
/db_xref="GI:7380063"
/db_xref="SPTREMBL:Q9JUB7"
/translation="MYRKTQAASRPPSAGDGAERAEHLNHNASNPILSPKRTVS"
/303409..303412
/303419..303697
/gene="NMA1405"
/303419..303697
/gene="NMA1405"
/note="NMA1405, possible periplasmic protein, len: 92 aa;
unknown, contains a possible N-terminal signal sequence"
/codon_start=1
/transl_table=1
/product="putative periplasmic protein"
/protein_id="CAB84645.1"
/db_xref="GI:7380064"
/db_xref="SPTREMBL:Q9JUB4"
/translation="MNIKTAFAICAIASAAAAVANEIKIDANNTPYSEADQKLA
TAVMGVKEPISLISGSGSITVSGSATQCVFKNQCALQIOGLNCK"
/303708..303740
/note="stem loop containing DNA uptake sequences: aaaa
gccgtctgaa gcc ttcagacgac atttt"
/303713..303722
/note="Core DNA uptake sequence: gccgtctgaa"
/Label=DUS
/misc_feature complement(303726..303735)
/note="Core DNA uptake sequence: gccgtctgaa"
/Label=DUS
/complement(303753..304370)

```

```

CDS
/gene="recR"
/complement(303753..304370)
/gene="recR"
/note="NMA1406, recR, possible recombination protein, len:
205 aa; similar to many e.g. SW:RER_ECOLI (EMBL:X15761),
recR, Escherichia coli recombination protein (201 aa),
fasta scores: E(): 6.9e-27, 42.4% identity in 191 aa
overlap"
/codon_start=1
/transl_table=1
/product="putative recombination protein"
/protein_id="CAB84646.1"
/db_xref="GI:7380065"
/db_xref="SPTREMBL:Q9JUB6"
/translation="MSHKODAFQGLIDALKVPWVGKSAQRIAYHLLQHKREPAEK
LYDALOTLAKQVDHCARCNTFEGGICIDICDETGGRLWVHPADVSNNEANCH
DGLYFVLMGOINTEAGMVSATALDRLOAGCGVEEIIATFATAGCNATAYVLS
PFKNLPYVSRSISQIPIAGGELEYDAGTLAQAVERRLKEGA"
/complement(304437..305975)
/gene="NMA1407"
/complement(304437..305975)
/gene="NMA1407"
/complement(304437..305975)
/gene="NMA1407"
/note="NMA1407, possible peptidyl-prolyl cis-trans
isomerase, len: 512 aa; similar to e.g. SW:CYPD_ECOLI
(EMBL:AE000150), ppib, Escherichia coli peptidyl-prolyl
cis-trans isomerase D (EC 5.2.1.8) (623 aa), fasta scores:
E(): 8.4e-14, 21.7% identity in 608 aa overlap. Contains a
probable N-terminal signal sequence"
/codon_start=1
/transl_table=1
/product="putative peptidyl-prolyl cis-trans isomerase"
/protein_id="CAB84647.1"
/db_xref="GI:7380066"
/db_xref="SPTREMBL:Q9JUB5"
/translation="MHSIEKTRPPAQLVLIATTPGEGSVYSHGADTYOVGD
EKISDHSINMAIIONEOADGGPSRDVAFOSLQRAYLKQAKLGI SVSSQIKQITV
DDPNFHDANGKEDHALNRLYSORHMSDOVEEIRDFALQNLVNLVONGVLDQAO
AEOLIRLTOVNRITRSHFNDEFAQVSEADQKRYNNKRDYLLPKAVKLEYVA
LNLKDFADKQTVSETEVNAFEERPARLPANDAKSPFOEKAAYAVNELMKKAVADFN
KAKKEIGDADFNHPSLSLAEAKNGLKQETWISRODAQMSGRENLIJAVSDDV
LKKHNSFVLITINSETAVYAKVEYREKTLPPFAKAKAVQATYRTAKALAEKAK
DLVTQNGKAVADVAKMSVSVLQAQAOQMSPEVATLLKAPANGKPAVYRLGLP
APVLTGVQAVTPPDIAQLPLAKQALAOQOSANTFDLLIRYFNKIKOTGAQSVDN
GDQ"
/complement(306029..306403)
/gene="NMA1408"
/complement(306029..306403)
/gene="NMA1408"
/note="NMA1408, len: 124 aa; unknown, similar to bacterial
hypothetical proteins e.g. SW:YFPB_ECOLI (EMBL:X57403),
yifB, Escherichia coli hypothetical protein (118 aa),
fasta scores: E(): 5.8e-15, 38.5% identity in 117 aa
overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1408"
/protein_id="CAB84648.1"
/db_xref="GI:7380067"
/db_xref="SPTREMBL:Q9JUB4"
/translation="MILHGIIPCDITYKAKNRLAGVGLFEFFDFKQRPSEAEICS
WLEQVPLATLNLKRGTSWRKIDAEQOQVLSSTAEAVYLNSEMSLTKRPVLECGKV
YASPEETVYDGIFFRQAPCRQG"
/complement(306287..306296)
/gene="NMA1408"
/note="Core DNA uptake sequence: gccgtctgaa"
/Label=DUS
/misc_feature complement(306474..306483)
/note="Core DNA uptake sequence: gccgtctgaa"
/Label=DUS
/gene="NMA1409"
/306553..308181
/306553..308181
CDS

```

```

A (P-loop)"
/codon_start=1
/transl_table=1
/product="putative ATP-dependent protease"
/protein_id="CAB84639.1"
/db_xref="GI:7380058"
/db_xref="SPTREMBL:O9JUC0"
/transl_table="MTQEKHFEEYALATLPRLDVVYPMVPLFVGRKSTAALE
NAITREPVFLAQTDAVEEPAADLYKOTVAVLOVLKLPDGTVALVEGYGR
VLTIEDGFLFIEAEVEDTGNDLDAVRRLLAOKFPOYAKINKKIPAEITIGSI
NGTANSRITDVTVAHQKIAQROQLIEPEIKRMFLLAQLESLDINQAKRIR
GRVKRQEKSRVAYLNQKAIHKEEGEDENGLDLEADIKKAGTKFAEKCUS
ELKTKMPPMSASTVARNYIDTLELPEMKRSVSKDAKALVDTDHYGLEKYK
ERILEYLAOKRMOKLGPITCLVGPVGKTSIGESIADKATGRKVMALGVDES
EIRGRRTYIGSMRKLQONMAKAGKNPLFLDEIDKSGDPGDPASALLEVDE
ONKFEADHYAEVDLSDVMFIATNSINIPILDMREITRLSGYDEDEKINIMQY
LYPKOMKNGCKESELVEESAVIDIIRYTRRAGVRSIDREIKICKKVMQITLDE
DKRRLSEKTKTSKAKPAVKVNEKLNHDYLGVRPFDGVAESNRIGQVGTLANTEVG
GELLTVEAALPGKGVIOCTGLDVMKESVSAAMSVRSRSEVGLASDFYEKRDH
VHVEGATPKDGPAGIAMTLAMVSAFSKIPRADVAMTGEITLRGEVLPIGKLEK
LAALRGSIKHVLIPEKDNKDLDEEIPENVKTGLTIHPVKMIDEVALGLSGPEPMAEP
SGAAVESAASKPKPRSRATKH"
/complement(295937..295963)
/gene="lon"
/note="P501046 ATP-dependent serine proteases, lon family,
serine active site"
/complement(296396..296980)
/gene="lon"
/note="Pfam match to entry PF00004 AAA, ATPases associated
with various cellular activities (AAA), score 113.90,
E-value 3e-30"
/complement(296942..296965)
/gene="lon"
/note="P500017 ATP/GTP-binding site motif A (P-loop)"
298136..298175
/note="Stem loop containing DNA uptake sequences: gccgtcgtgaa
gccgtcgtgaa acattccggc ttccagcggc atggg"
298141..298150
/note="Core DNA uptake sequence: gccgtcgtgaa"
/label=DUS
/complement(298161..298170)
/note="Core DNA uptake sequence: gccgtcgtgaa"
/label=DUS
/complement(298398..299393)
/label=DUS
/gene="NMA1400"
298398..299393
/gene="NMA1400"
298398..299393
/gene="NMA1400"
/note="NMA1400, len: 331 aa: unknown. shows weak
similarity to parts bacterial hypothetical proteins e.g.
TR:CAB49354 (EMBL:AJ248284) Pyrococcus abyssi hypothetical
protein (197 aa), fasta scores: E(): 1.5e-06, 33.5%
identity in 179 aa overlap. Contains P500092 N-6
Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1400"
/protein_id="CAB84640.1"
/db_xref="GI:7380059"
/db_xref="SPTREMBL:O9JUB9"
/transl_table="MKATONIAITWGDHNNAKOVL SAMKKYKRPAAVRSADIA
LFNAHNRKQAOOSRLINMLAVEIRPGVLDNKRAPDIRAALLDVGADKRPFLPIN
LLLFQCAHEMHKGVAVPOLGSIHVPGFVSLRGEYLDLALAHASFOTAFDIDG
TGGSVLAALAKOGIPSVIGTDNPKAVACARANIALRGEKOVEIRETDLFPGGFAD
LIVCNPMFLPAKPTSAVESALYDPESAMLAFLDAPKHLNPDGEITRLIISDLAEHLH
LRSPDLKAPAGAGLPAVDVLTCTKPYHKKAADPNPLAARNRETFTLYRKKKA"
/complement(298855..298864)
/note="Core DNA uptake sequence: gccgtcgtgaa"
/label=DUS
299055..299075
/gene="NMA1400"
/note="P500092 N-6 Adenine-specific DNA methylases
signature"
/complement(299465..301210)

```

```

CDS
/gene="recD"
/complement(299465..301210)
/gene="recD"
/note="NMA1401, recD, exodeoxyribonuclease V alpha
subunit, len: 581 aa: highly similar to TR:068924
(EMBL:AF058330), recD, Neisseria gonorrhoeae
exodeoxyribonuclease V alpha subunit (581 aa), fasta
scores: E(): 0, 95.2% identity in 580 aa overlap. Contains
P500017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=1
/product="exodeoxyribonuclease V alpha subunit"
/protein_id="CAB84641.1"
/db_xref="GI:7380060"
/db_xref="SPTREMBL:O9JUB8"
/transl_table="MEROTDEFAQAAARAIIRLEHYVSSGDEVLANCTERFOLON
GHSFIRLSGGRADLSALAPVGTSAAPLIRLEGRPLIGRMWOLEFYDAAEIKRLAA
DTSAPDAAGARONLAKMFQGTSGEQDAALALLOSFMVITGGGPGTKTTVAKLLA
LICKEDENLPILALAPAKGAAMARALHRAINGFDALAEVRRHLLEGGTVARHL
KLRRPKMOAENPYPLPEVDLVIDEASMLDTALMOLKAKVKGARVITLIGENOLP
NSVIGAVLSVLSOKTVLDGETHORLAGLPEHFGSVSNPVPVLAQNTAHLSESRFGD
NSGICGLARAASGDEGAMALFDRRPDLTSHSGSPNARVRLTRAHKAYQADKGN
IEAAYAGISDILVLAAMRQDNEDNEAYCRVRRKMNIPERLAFPGPPIRQNDTA
LELFNGDIGLMEVDYGRQSLAAFPADGKRKAVSCLPEEPFAFTVHRSQSEY
REWLLPSPASPSDEGDALSGLSKELYTAITRAREKVFEGGEETFRQAAATVTR
QTALGSMLERYFSOG"
/complement(300755..300778)
/gene="recD"
/note="P500017 ATP/GTP-binding site motif A (P-loop)"
/complement(301216..301219)
301224..301262
/note="Stem loop containing DNA uptake sequences: caaat
gccgtcgtgaa gccgtcgtgaa ttccagcggc gcgg"
301229..301238
/note="Core DNA uptake sequence: gccgtcgtgaa"
/label=DUS
/complement(301249..301258)
/note="Core DNA uptake sequence: gccgtcgtgaa"
/label=DUS
/complement(301278..301973)
/gene="NMA1402"
/complement(301278..301973)
/gene="NMA1402"
/note="NMA1402, probable ABC-transporter ATP-binding
protein, len: 231 aa: similar to many e.g. SW:GINO.BACST
(EMBL:M61017), glnQ, Bacillus stearothermophilus glutamine
transport ATP-binding protein (242 aa), blast scores,
Expect 3.6e-29 and SW:YCEV.ECOLI (EMBL:AE000212), yciY,
Escherichia coli hypothetical ABC transporter ATP-binding
protein (233 aa), fasta scores: E(): 0, 53.8% identity in
223 aa overlap. Contains Pfam match to entry PF00005
ABC_tran, ABC transporter, P500211 ABC transporters family
signature and P500017 ATP/GTP-binding site motif A
(P-loop)"
/codon_start=1
/transl_table=1
/product="putative ABC-transporter ATP-binding protein"
/protein_id="CAB84642.1"
/db_xref="GI:7380061"
/db_xref="SWISS-PROT:P57030"
/transl_table="MSLLILKCEGVGRYRQGLDVLVNLGLDEIRAGSTGIIGSS
GASGKSTLHLIGGIDMSSEGLVLDGSLRNLNRGLDNRNLGIVYOFHLLPEF
SLSENWVWPLIGIKSSSEEAAMAMLEVKYHRTSHAGSELGSGEORAAIARAL
VTQPKCLADEPNCILDRANARVLDMMLEKLTGLGVVYVHDDLAORPERVYVM
KDSLIHFKQGANR"
/complement(301311..301871)
/gene="NMA1402"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 209.80, E-value 4e-59"
/complement(301494..301538)
/gene="NMA1402"
/note="P500211 ABC transporters family signature"
/complement(301785..301794)

```

```

misc_feature
/label=DUS
/misc_feature
299055..299075
/gene="NMA1400"
/note="P500092 N-6 Adenine-specific DNA methylases
signature"
complement(299465..301210)

```

```

misc_feature
/misc_feature
misc_feature

```

```

/misc_feature /note="Core DNA uptake sequence: gccgctctgaa"
/label-DUS
complement(291221..291230)
/gene="NMA1393"
/note="Core DNA uptake sequence: gccgctctgaa"
/label-DUS
complement(291229..291645)
/gene="NMA1393"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 36.50, E-value 3.3e-08"
complement(291403..291447)
/gene="NMA1393"
/note="PS00211 ABC transporters family signature"
complement(291669..291676)
/note="(cg)4"
complement(292023..292610)
/gene="NMA1393"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 150.70, E-value 2.5e-41"
complement(292197..292241)
/gene="NMA1393"
/note="PS00211 ABC transporters family signature"
292399..292408
/note="Core DNA uptake sequence: gccgctctgaa"
/label-DUS
292541..292631
/note="termination of replication?"
/label-terminator?
complement(292566..292589)
/gene="NMA1393"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
292746..292755
/note="Core DNA uptake sequence: gccgctctgaa"
/label-DUS
complement(292765..293160)
/gene="NMA1394"
complement(292765..293160)
/gene="NMA1394"
/note="NMA1394, len: 131 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1394"
/protein_id="CAB84635.1"
/db_xref="GI:7380054"
/db_xref="SPTREMBL:Q9JUC2"
/transl_table=11
/product="putative homoserine dehydrogenase"
KDSVCLISOCRTLSDLTPENLISDIKPVDPYPELLIGTGAQEFHPKIMADPSRIGIS
VECMNTDSAFRLVFLHSEGRRAWMLQP"
complement(293234..293243)
/note="Core DNA uptake sequence: gccgctctgaa"
/label-DUS
293323..293327
293333..294640
/gene="hom"
293333..294640
/gene="hom"
/EC_number="1.1.1.3"
/note="NMA1395, hom, probable homoserine dehydrogenase,
len: 435 aa; similar to many e.g. SW:DHOM.PSPAE
(EMBL:X65033), hom, Pseudomonas aeruginosa homoserine
dehydrogenase (EC 1.1.1.3) (439 aa), fasta scores: E(): 0,
50.78 identity in 442 aa overlap. Contains Pfam match to
entry PF00742 Homoserine_dh, Homoserine dehydrogenase and
PS01042 Homoserine dehydrogenase signature"
/codon_start=1
/transl_table=11
/product="putative homoserine dehydrogenase"
/protein_id="CAB84636.1"
/db_xref="GI:7380055"
/db_xref="SPTREMBL:Q9JUR84"
/transl_table=11
/product="putative homoserine dehydrogenase"
LSEKAROTCSAAFKDPFELVAREDDVAVLELGGTGIADAVLKAENGKHVTA
NKILAEYGNELFLAEKQNVIVQFEAAVAGGIPILIKALREGIAANRIKSIAGIINGT

```

```

/misc_feature /note="Pfam match to entry PF00742 Homoserine_dh,
Homoserine dehydrogenase, score 484.20, E-value 1e-141"
293876..293944
/gene="hom"
/note="PS01042 Homoserine dehydrogenase signature"
294633..295067
/gene="NMA1396"
294633..295067
/gene="NMA1396"
/note="NMA1396, possible membrane protein, len: 144 aa;
unknown, contains a hydrophobic, possible
membrane-spanning region"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84637.1"
/db_xref="GI:7380056"
/db_xref="SPTREMBL:Q9JUC1"
/transl_table=11
/product="putative membrane protein"
SOCANSKPOAKOIVESCVCNIPFAEKQNDLRAKGLDSNTRPLTVDYCKCMWQPLD
RLSEKOISSFGGLAQEOEDLLGANNARETRDKQCVADLKE"
295079..295112
complement(295126..295395)
/gene="hupB"
complement(295126..295395)
/gene="hupB"
/note="NMA1397, hupB, probable DNA-binding protein, len:
89 aa; similar to many e.g. SW:DBHB.ECOLI (EMBL:X53241),
hupB, Escherichia coli DNA-binding protein HU-beta (90
aa), fasta scores: E(): 1.4e-19, 60.5% identity in 89 aa
overlap. Similar to NMA1868, fasta scores: E(): 3.6e-21,
69.7% identity in 89 aa overlap. Contains Pfam match to
entry PF00216 Bac_DNA-binding, Bacterial DNA-binding
protein and PS00045 Bacterial histone-like DNA-binding
proteins signature"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="CAB84638.1"
/db_xref="GI:7380057"
/db_xref="SPTREMBL:Q9JRI6"
/transl_table=11
/product="putative DNA-binding protein"
VGFGEYVGERERGRNPKTGEPLTIAAARTPKRAGKALKDAL"
complement(295129..295395)
/gene="hupB"
/note="Pfam match to entry PF00216 Bac_DNA-binding,
Bacterial DNA-binding protein, score 143.60, E-value
3.4e-40"
complement(295201..295260)
/gene="hupB"
/note="PS00045 Bacterial histone-like DNA-binding proteins
signature"
complement(295405..295410)
complement(295577..298039)
/gene="lon"
complement(295577..298039)
/gene="lon"
/EC_number="3.4.21.53"
/note="NMA1398, lon, probable ATP-dependent protease, len:
820 aa; similar to many e.g. SW:LON.ECOLI (EMBL:L12349),
lon, Escherichia coli ATP-dependent protease (EC
3.4.21.53) (784 aa), fasta scores: E(): 0, 65.5% identity
in 796 aa overlap. Contains Pfam match to entry PF00004
AAA, Aarases associated with various cellular activities
(AAA), PS01046 ATP-dependent serine proteases, lon family,
serine active site and PS00017 ATP/GTP-binding site motif

```



```

RBS      /note="Pfam match to entry PF01334 Bacteriofer,
gene      Bacterioferritin, score 345.90, E-value 4.3e-100"
CDS      complement(280390..280394)
          /gene="lfpB"
          /note="lfpB"
          /complement(280412..280876)
          /gene="bfrA"
          /note="bfrA"
          /complement(280412..280876)
          /gene="NMA1377, bfrA, probable bacterioferritin A, len:
          154 aa: similar to many e.g. SW:BRA_NEICO (EMBL:U07653),
          bfrA, Neisseria gonorrhoeae bacterioferritin A (154 aa),
          fasta scores: E(): 0, 99.4% identity in 154 aa overlap and
          SW:BFR_PSEPU (EMBL:U06717), bfr, Pseudomonas putida
          bacterioferritin (154 aa), fasta scores: E(): 0, 61.0%
          identity in 154 aa overlap. Similar to NMA1376, fasta
          scores: E(): 2.2e-17, 39.9% identity in 153 aa overlap"
          /codon_start=1
          /transl_table=1
          /product="putative bacterioferritin A"
          /protein_id="CAB84623.1"
          /db_xref="GI:7380045"
          /db_xref="SWISS-PROT:P56998"
          /translation="MGSQAVVDYMNELSGELARDQYFIHSRLYSEMGYTKLPERL
          NHEEETTHAEDFIRILMIGTPKMARALNIGTGVSCAKADLTQTEYEVRLALK
          GIKLCEADQYVSRDLVAQLKDTEDHAHWLEQQLRIELIGBNYYQSQL"
          /complement(280415..280876)
          /gene="bfrA"
          /note="Pfam match to entry PF01334 Bacteriofer,
          Bacterioferritin, score 360.40, E-value 1.9e-104"
          /complement(280882..280887)
          /complement(281249..281258)
          /note="Core DNA uptake sequence: gccgctcgaa"
          /label=DUS
          /complement(281340..282323)
          /gene="lfpA"
          /complement(281340..282323)
          /gene="lfpA"
          /complement(281340..282323)
          /note="NMA1378, lfpA, probable lipoteic acid synthetase,
          len: 327 aa: similar to many e.g. SW:LIPA_ECOLI
          (EMBL:U07636), lfpA, Escherichia coli lipoteic acid
          synthetase (321 aa), fasta scores: E(): 0, 63.3% identity
          in 308 aa overlap"
          /codon_start=1
          /transl_table=1
          /product="putative lipoteic acid synthetase"
          /protein_id="CAB84624.1"
          /db_xref="GI:7380046"
          /db_xref="SPTREMBL:09JUC8"
          /translation="MSEIKTDDPKRGIKLGADKTAIRPIKVVPLQEKLKPEMIRAK
          LPSRKFEIIDLREQMHTVCEASCPNIGECFSKGTAFPMJMGDICTRCEPCDVG
          HGRPNMLDPDEPKRLAESVKAMNRYVITSVDRLDROGAHPADCIKAIARETSPN
          TKIILVDPGRGLDIAKILAEIPDVMMNHL.ETHPSLYRKARPCANYSHSDILKR
          YKEMPHIPKSGIMVIGICTEDDVREIMDMRAHNEMITTCOTYLOPSGHLPLEKY
          VPEQGIKPEKVEYELGFSNMAIGAMVRSYHDEQNAELRSHGCGHH"
          /complement(282316..282888)
          /gene="lfpB"
          /complement(282316..282888)
          /gene="lfpB"
          /complement(282316..282888)
          /EC_number="6.-.-.-"
          /note="NMA1379, lfpB, probable lipoteic acid synthetase,
          len: 190 aa: similar to many e.g. SW:LIPB_ECOLI
          (EMBL:U07636), lfpB, Escherichia coli lipoteic acid
          synthetase (191 aa), fasta scores: E(): 0,
          57.9% identity in 178 aa overlap. Contains P501313
          lipoteic acid synthetase B signature"
          /codon_start=1
          /transl_table=1
          /product="putative lipoteic acid synthetase B"
          /protein_id="CAB84625.1"
          /db_xref="GI:7380047"
          /db_xref="SPTREMBL:09JUC7"
          /translation="MKTNAGRAADTDEDELAVVHPVFTQGLVGRPEHLIRDDIPV
          VOIDRGQITVHPQGLVVTMDIFRRRTSVNIVSALENSIATVLAEGIEAADP
          KRGVYVGRKIASIGLRIRKNSVYHGLALNVMDLSPFTHINPCGACMEMTQIADP

```

```

misc_feature VOPCPDPDEVAKLTALHETPTPRADNNE"
complement(282697..282744)
/gene="lfpB"
/note="PS01313 Lipoteic acid synthetase B signature"
complement(282941..283216)
/gene="NMA1380"
complement(282941..283216)
/gene="NMA1380"
complement(282941..283216)
/gene="NMA1380"
complement(282941..283216)
/note="NMA1380, len: 91 aa: unknown, similar to bacterial
hypothetical proteins e.g. SW:YBED_ECOLI (EMBL:U07636),
ybed, Escherichia coli hypothetical protein (87 aa),
fasta scores: E(): 3.1e-08, 37.8% identity in 82 aa
overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1380"
/protein_id="CAB84626.1"
/db_xref="GI:7380048"
/db_xref="SPTREMBL:09JUC4"
/translation="MTEDKNTSLIEPCGTPPLKVMGAVHPEPQAVLDIVRLHAPPY
QAHITTRPSSKGNVTGATVQVKNEDQDLNIVRALTSHELVKVL"
283239..283281
/note="Stem loop containing DNA uptake sequences: ggat
gccgctcgaa gccgaggttgcct ttcaagcggc atcgg"
283244..283253
/misc_feature /note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
complement(283267..283276)
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
283401..284471
/gene="NMA1381"
283401..284471
/gene="NMA1381"
/note="NMA1381, probable integral membrane protein, len:
356 aa: similar to many bacterial hypothetical proteins
e.g. SW:PEBM_ECOLI (EMBL:AB000335), perm, Escherichia coli
putative permease (353 aa), fasta scores: E(): 1.5e-19,
26.2% identity in 313 aa overlap. Contains Pfam match to
entry PF01594 DUF20, Putative permease. Contains
hydrophobic, probable membrane-spanning regions"
/codon_start=1
/transl_table=1
/product="putative integral membrane protein"
/protein_id="CAB84627.1"
/db_xref="GI:7380049"
/db_xref="SPTREMBL:09JUC6"
/translation="MYRRKRGKIKPMDACAFAALVPLVPAUGDITTPPAVAAYLAV
VDLPVYEMLOKRGKINRASMSVAVSLLILALDLITVPLVGOENNLASRLPQILG
FMONTLLPWLKNTIGTGVETDOASITAMLOAHGELSNAKAMPVYLMRGQNVISI
GNLLLPPLLKYPLLDQWRSCGIAKLVPRFAGATYRTIGNLNVEYGEFLRGQLLVM
LNLVYVGLVGLVGLDSGFAIGMVGAILVFEVYLGAFTGLLATVAAALQFSSNGI
LAVMAVAVGQFLSEFPTKIVGDRIGLSPFWIVISIMAFGLMVGMLADLPPLAA
VTLVLLREGVQKVFASGFTYGR"
283437..284432
/gene="NMA1381"
/note="Pfam match to entry PF01594 DUF20, , score 180.90,
E-value 2.1e-50"
284524..284568
/note="Stem loop containing DNA uptake sequences: aaaa
gccgctcgaa cccgaaacgggac ttcaagcggc atttt"
284529..284538
/misc_feature /note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
complement(284554..284563)
/note="Core DNA uptake sequence: gccgctcgaa"
/label=DUS
complement(284578..285525)
/gene="NMA1382"
complement(284578..285525)
/gene="NMA1382"
/note="NMA1382, possible periplasmic protein, len: 315 aa:
similar to many hypothetical proteins found in both

```

```

/misc_feature /db_xref="SPTREMBL:Q9JUD0"
/translacion="KRIYEKATIEDVDLVLPVASHSVLEPRDYKLOTKLTRETTLNPLLSAMDTVEARLAISMAOEGGIGITIHKNMPPEMOKARLSKVRHESGVXDPVPVAPFTLIREVLEAKORRRKMSGLPVENGGKVGIVTNDLFEFNRVDLPVSAIMTPERELVTYPETSIDEARLMTHTKHVERVLNLNKDELKITVKDLKTTEFPNAKDSEGRLLRYGAVALDKEEERKALVEGVYIVPDVPHSGGVLDIRVMWVEYTHIOVIGGILNATPKALLDIYAAGADAVKVGIGGISTCTTRIYACVGPOLPTLIHNVAALKTGPPILIADGIRPSGDIAKALAAGASVWLGMGFTEBAPEIELYIGRSTKSTRNGSLGAAVSQADRYFDCKTDSDAKTYPEGISERVYKGPYIVIHQTLGRLSSNGYLGCANNIAEMGEAPEVEITSGASESHVDVOTKTEAPNYHR"
/gene="guab"
/complement(274408 . 275076)
/note="Pfam match to entry PF00478 IMPDH_C, IMP dehydrogenase / GMP reductase, score 432.80, E-value 3e-126"
misc_feature complement(274861 . 274899)
/gene="guab"
/note="PF00487 IMP dehydrogenase / GMP reductase signature"
complement(275164 . 275325)
/gene="guab"
/note="Pfam match to entry PF00571 CBS, CBS domain, score 59.80, E-value 6e-14"
complement(275350 . 275514)
/gene="guab"
/note="Pfam match to entry PF00571 CBS, CBS domain, score 37.30, E-value 3.6e-07"
complement(275518 . 275781)
/gene="guab"
/note="Pfam match to entry PF01574 IMPDH_N, MP dehydrogenase / GMP reductase C terminus, score 166.40, E-value 4.7e-52"
complement(276108 . 276569)
/gene="NMA1373"
complement(276108 . 276569)
/gene="NMA1373"
/note="NMA1373, len: 153 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1373"
/protein_id="CAB84619.1"
/db_xref="GI:7380041"
/db_xref="SPTREMBL:Q9JQM2"
/translacion="NIITMDSTRKRGILMEBNSCNCFYKIQIITIGENIYPIKICPYGTFLINAFVNLSTSFEEERYAGGNNGNDLFDEQLFDIDKYNLSLCLMFSDITYMSGGNCNCIDICLVLMGSGEEERLYFSFDNGKNFKETIRYKGTVESVIFQLNL"
complement(276575 . 276578)
complement(276612 . 276621)
/note="Core DNA uptake sequence: gccgtctgaa"
misc_feature label=DUS
276651
276642 . 276651
/note="Core DNA uptake sequence: gccgtctgaa"
label=DUS
complement(276660 . 279218)
/gene="glnb"
complement(276660 . 279218)
/gene="glnb"
/gene="glnb"
/ec_number="2.7.7.59"
/note="NMA1374, glnb, probable [protein-pii] uridylyltransferase, len: 852 aa; similar to many e.g. SW:GIND_ECOLI1 (EMBL:M96431), glnb, Escherichia coli [protein-pii] uridylyltransferase (EC 2.7.7.59) (890 aa), fasta scores: E(): 0, 33.1% identity in 865 aa overlap"
/codon_table=1
/transl_table=11
/product="putative [protein-pii] uridylyltransferase"
/protein_id="CAB84620.1"
/db_xref="GI:7380042"
/db_xref="SPTREMBL:Q9JUC9"
/translacion="MPANLSSALFFEPKOORDAFAAHYIKANRVSPPRETAVERLLAALNAEYONSAKLCLMAVGGFRGBELLPCSDVDLAYVSPAPLSDGIDEQIARVOTIMDCKLIPVSKSSVSELCEVSNVIDTGPALFEARLFENROQTADKLAEKNAQSYOAEFEALVIEPVRHHKSGSGAVILEPNITKSCGGGLRDITHTLMLTAKAQGLTDDLPDLK
```

misc_feature
278198..278207
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS

misc_feature
278952..278961
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS

gene
complement(279267..279587)
/gene="NMA1375"

CDS
complement(279267..279587)
/gene="NMA1375"
/note="NMA1375", possible transcriptional regulator, len: 106 aa; similar to e.g. regulatory proteins of restriction modification systems e.g. Rf:Q52622 (EMBL:M77223), puvrIC, Proteus vulgaris regulatory transcription factor C (84 aa), fasta scores: E(): 1.8e-05, 42.2% identity in 64 aa overlap. Similar to TR:AD03136 (EMBL:Arl23569), tmeSIC, Neisseria meningitidis strain 800615 (serotype B:d:P1.4) putative control element in 70 aa, fasta scores: E(): 4.3e-05, 46.0% identity in 50 aa overlap. Contains Pfam match to entry PF01381 HPH_3, Helix-turn-helix. Contains a probable helix-turn-helix motif at aa 36-57 (Score 1818, +5.38 SD).
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="CAB84621.1"
/db_xref="GI:7380043"
/db_xref="SPTRMBL:OJ9JRO9"
/translation="MNGKKTLPLALPEPDQLRAVLVYNNRFLPVNKGNSOEELARCG GIDRVYSAVEKRRNMIALSNTEKMAALGVAAIOLPLPOERKLTKTNSDTRÖMPS ESGI"

misc_feature
complement(279279..279288)
/gene="NMA1375"

misc_feature
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS

misc_feature
complement(279348..279512)
/gene="NMA1375"
/note="Pfam match to entry PF01381 HPH_3, Helix-turn-helix, score 49.60, E-value 6.9e-11"
complement(279911..280384)
/gene="bfirB"

gene
complement(279911..280384)
/gene="bfirB"

CDS
complement(279911..280384)
/gene="bfirB"
/note="NMA1376, bfirB, probable bacterioferritin B, len: 157 aa; similar to many e.g. SW:BFRB_NEICO (EMBL:U76634), bfrB, Neisseria gonorrhoeae bacterioferritin B (157 aa), fasta scores: E(): 0.994% identity in 157 aa overlap and SW:BFR_AZOV1 (EMBL:M83692), bfr, Azobacter vinelandii bacterioferritin (156 aa), fasta scores: E(): 0.622% identity in 156 aa overlap. Similar to NMA1377, fasta scores: E(): 3.1e-17, 39.9% identity in 153 aa overlap"
/codon_start=1
/transl_table=1
/product="putative bacterioferritin B"
/protein_id="CAB84622.1"
/db_xref="GI:7380044"
/db_xref="SWISS-PROT:P56599"
/translation="MKGDRLVIRELNKLGILLVTINQYFHLRIILKNMGFEELGHEFF FKQSLVEAKADDLIERLFLEGLPIPNDELGLLTIGSTETIEIACDTIKRKDEHALL AAIATAEKOQDVSRDLLEKQDINEEHITWLETFQEGISGLGPNTLTAAQED" complement(279923..280384)
/gene="bfirB"

misc_feature
complement(279923..280384)
/gene="bfirB"

| | |
|--------------|---|
| misc_feature | /note="Coreia element; hmms hit to HM Correira (1 - 156), score: 270.84" /label=Correira complement(268713..268722) /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS 268803..270614 /gene="typha" 268803..270614 /gene="typha" /note="NMA1370, typhA, probable GTP-binding protein, len: 603 aa; similar to many e.g. SW:TYPA_ECOLI (EMBL:J19201), typhA, Escherichia coli GTP-binding protein implicated in regulation of virulence factors, fasta scores: E(): 0, 64.1% identity in 599 aa overlap. Contains Pfam match to PF00301 GTP-binding elongation factor Tu family, PF00017 ATP/GTP-binding site motif A (P-loop)" /codon_start=1 /transl_table=11 /product="putative GTP-binding protein" /protein_id="CAB84616.1" /db_xref="GI:7380038" /db_xref="SP:REMBL:09JUD2" /translation="MKQIRNTAIIAHVHGKTTLVLDQILROSGTFRANQOQDERVWDS NDLERKSTITLAKNTAIDESTHINIVDTPEHAFGEVEERVLGMCVLLVAQAE GPMPOTRVTYKALALGLKPIYVINKIDKPSAPSMVIDQEPFLDNGATGLODFP IYVAGLSGFAKLEETDESNDMPLEDTLTKTPAPSGADETLOLOLSQIDYNTG RLGIRIINGKIKPGQVAVMNHDOQIAGRIKNOILGFKGLERVLLEERAGDIIYIS GIEDIGVITTDKNDKPGIPLMSYDEPTLTHDPVNTSPLAGTGKRVTSQIDRL OKELITVALREVEDADADPRVSRGELHLITLLENNRREGYELAVGKPRVYDID GCKEPEYNLVDPDNDQAGVAMELGRRRELITMEDSGRGTLEVEYHIPRGDIGE OGERHTLRGLMSHVEDDAPVAPKPMGPRHNGVLSQEQEAVAVALMLEDGRM FESPDKTYEEMIGIHSRDNDLVNPLKGRKLTIRASGDEAVRLTPIKLTIEGA VFETIDDELVEITPQISIRLRKRYSELERRRHFKKD" |
| misc_feature | /gene="typhA" /note="Pfam match to entry PF00009 GTP_EFTU, Elongation factor Tu family, score 354.60, E-value 1.1e-102" 268833..268856 /gene="typhA" /note="PS00017 ATP/GTP-binding site motif A (P-loop)" 268929..268976 /gene="typhA" /note="PS00301 GTP-binding elongation factors signature" 270644..270677 /note="Stem loop containing DNA uptake sequences: gccgtctgaa aaat ttcaagcagc atttt" 270649..270658 /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(270663..270672) /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(270713..270722) /note="Core DNA uptake sequence: gccgtctgaa" 270876..271030 /note="Correira element; hmms hit to HM Correira (1 - 156), score: 274.61" /label=Correira complement(271178..271268) /note="tRNA Ser anticodon GGA, Cove score 68.79" complement(271307..271393) /note="tRNA Leu anticodon CAG, Cove score 72.49" complement(271429..271515) /note="tRNA Leu anticodon CAG, Cove score 71.18" 271604..271979 /gene="vacB" 271604..271979 /gene="vacB" /note="NMA1371, vacB, probable ribonuclease, len: 791 aa; similar to many e.g. SW:VACB_SHIFL (EMBL:D11024), vacB, |
| misc_feature | Shigella flexneri chromosomal gene required for expression of virulence genes (also in enteroinvasive Escherichia coli) (760 aa), fasta scores: E(): 0, 44.9% identity in 762 aa overlap. May have a post-transcriptional regulatory role. Also similar to SW:RNB_ECOLI (EMBL:X67913), rnb, Escherichia coli exoribonuclease II (EC 3.1.13.1) (644 aa), fasta scores: E(): 0, 30.0% identity in 580 aa overlap. Contains Pfam matches to entry PF00773 RNB, RNB-like proteins and entry PF00575 S1, S1 RNA binding domain and PS01175 Ribonuclease II family signature" /codon_start=1 /transl_table=11 /product="putative ribonuclease" /protein_id="CAB84617.1" /db_xref="GI:7380039" /db_xref="SP:REMBL:09JUD1" /translation="MNKNIKSLMRKEDPLSEKORYEHLPSEWMIITELRKGVPSKEALVRELSTEEREEFFERLRKAMARGOVLINRGAACAADKIDLVKREAKR DCGEFLVPLTPADGDFVLYERQRMIGMDITVPARGDRREGVLDIVERO SKVYGEYMDRGVALIEPEKRLNOSIVLEDPKSGVIGVETIEVPEQNP AAKIIEVLDGYADSGMEIEIAVRKHLHPQSEACAKAKKIPHYRKSGLGRDVL RDLPLVITIDGETARDPDAVFAEKGIRNRYIAYIADSHYRPDAIDIDAOERST VYEPRIIVLPEPNSINGCISLNPHYERLCVQDMYITAGNIKEYRFPVAMRSAR LTYNWMKLSGIEHPKTYQIDTLTKLERLILQKRFERGAVEFSDIETOMLEDNGK IETKYVYVNDHAKLIEECMLANWCAADPLKNTALFRNHLGTPPKLAAREOL GLGLQVLDGGDNPSPDYALAGQFGRPDAELQVMMLRMQQAVYEPHCGHPELA YEAVHFTSPIRRYDPLVYHRAIKAVLNQOYTPPKSMQALCVHSTSCRRADASRD VENMLKTYRMKRYKGVPEGRKISGMTSPFITPTDGIHIDGLVHISDLGDFENFPE IMAIGERSGIRFENMGDRVAAYARADLDGKIDIVLLAGSGGRKRVSSASAPAG TAGKRPKTAARKKIRARGKIVGRASAAEERAKKRPVPIKKRKKS" |
| misc_feature | /gene="vacB" /note="Pfam match to entry PF00773 RNB, RNB-like proteins, score 506.80, E-value 1.6e-148" 273287..273361 /gene="vacB" /note="PS01175 Ribonuclease II family signature" 273506..273763 /gene="vacB" /note="Pfam match to entry PF00575 S1, S1 RNA binding domain, score 51.30, E-value 2.1e-11" complement(274160..274169) /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS 274252..274296 /note="Stem loop containing DNA uptake sequences: aaat gccgtctgaa acccgaaatcaggt ttcaagcagc atttt" 274257..274266 /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(274282..274291) /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(274318..275781) /gene="guab" complement(274318..275781) /gene="guab" /EC_number="1.1.1.205" /note="NMA1372, guab, probable inosine-5'-monophosphate dehydrogenase, len: 487 aa; similar to many e.g. SW:IMDH_ECOLI (EMBL:AE000337), guab, Escherichia coli inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205) (488 aa), fasta scores: E(): 0, 64.4% identity in 491 aa overlap. Contains Pfam matches to entry PF00478 IMPDH_C, IMP dehydrogenase / GMP reductase, to entry PF01574 IMPDH_N, MP dehydrogenase / GMP reductase C terminus, two to to entry PF00571 CBS, CBS domain and PS00487 IMP dehydrogenase / GMP reductase signature" /codon_start=1 /transl_table=11 /product="putative inosine-5'-monophosphate dehydrogenase" /protein_id="CAB84618.1" /db_xref="GI:7380040" |

```

misc_feature
/complement(265339. .265377)
/gene="cysg"
/note="Pfam match to entry PF00590 TP-methylase,
necaravirrole (Corrin/porphyrin) Methylases.", score 286.40,
E-value 3.7e-82"

misc_feature
/complement(265636. .265737)
/gene="cysg"
/note="P500840 Uroporphyrin-III C-methyltransferase
signature 2"
265861. .265870
/note="Core DNA uptake sequence: gccgcctctgaa"

misc_feature
/label=DUS
/complement(265918. .265962)
/gene="cysg"
/note="P500839 Uroporphyrin-III C-methyltransferase
signature 1"
266637)
/complement(266899. .266907)
/note="Core DNA uptake sequence: gccgcctctgaa"

misc_feature
/label=DUS
/complement(266922. .267578)
/gene="NMA1368"
/complement(268922. .267578)
/gene="NMA1368"
/note="NMA1368, possible membrane protein, len: 218 aa;
similar to TR:Q6471 (EMBL:U25811) Chromatium vinosum
hypothetical protein (231 aa), fasta scores: E(): 1.1e-23,
37.9% identity in 211 aa overlap. Contains hydrophobic,
possible membrane-spanning regions. C-terminus contains a
poly-Ala tract"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB84614.1"
/db_xref="GI:7380036"
/db_xref="SPTREMBL:O9JUD4"
/transl_table="MKNRTKYMDEPTRLFWHLAASLPFMYSAKTGGMDLQNHTRVG
LTLAFLVFLRGLCWGIGSDTARFSRFVGSIGREYKNGIPENVOGPHGALMW
ALFAAATVFGTGLFRADENTFTSTNGYNHLYSETHGSLMKILHNFKLAIFSVAH
IAAANAAYRVFKRNLVLPMTIGFKYIEKRTSIRAGKAALAAALSVAAALAAAILLS
"

RBS
267634. .267638
267643. .268533
/gene="NMA1369"
267643. .268533
/gene="NMA1369"
/note="NMA1369, len: 296 aa; unknown, similar to the
C-terminal half of bacterial hypothetical proteins e.g.
SW:JFE-ECOLI (EMBL:U14003), yJEF, Escherichia coli
hypothetical protein (515 aa), fasta scores: E(): 7.8e-28,
37.8% identity in 278 aa overlap. Contains Pfam match to
entry PF01256 UPF0031, Uncharacterized protein family
UPF0031"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1369"
/protein_id="CAB84615.1"
/db_xref="GI:7380037"
/db_xref="SPTREMBL:O9JUD3"
/transl_table="MFPVFNHLSGGSRRRMLQALRPVHYFKARAEDSHKTFGLTAVY
GGAVMSAGSAPVLASAAAMYIISCGGVKMGFQNTLPFVAVIAGPEIMLDTADGLTRQK
INAMTVAGGGLSTDKRAAGTTLAGITTEHTDQKPVLLDADALNISTDAETFNLRGCKNI
LITPPASARARLGLTVAQVADRTAAYVRCATGATVYLGKHTLVASDTEIYVI
EGSGNGLTAGSGDVLGIIIGSLIAGVPPFEACAGAMLHGAAADVIKESGIAG
TAGETIAPARWLNRIRTESM"

misc_feature
267763. .268494
/gene="NMA1369"
/note="Pfam match to entry PF01256 UPF0031,
uncharacterized protein family UPF0031, score 169.90,
E-value 4.1e-47"
/complement(268539. .268693)

repeat_unit

```

repeat_unit

```

gene      /note="Stem loop containing DNA uptake sequences: cctat
          accgtctgaagacagagcgctgt ttcagacgct atcgaa"
          complement(258522..260291)
          /gene="cysJ"
          complement(258522..260291)
          /gene="cysJ"
          /EC_number="1.8.1.2"
          /note="NMA1362, cysJ, probable sulphite reductase beta
          subunit, len: 589 aa; similar to many e.g. SW:CYSJ_ECOLI
          (EMBL:U029579), cysJ, Escherichia coli sulfite reductase
          [NADPH] hemoprotein beta-component (EC 1.8.1.2) (569 aa),
          fasta scores: E(): 0, 68.7% identity in 566 aa overlap.
          Contains Pfam match to entry PF01077 NIR_SIR, Nitrite and
          sulphite reductase and PS00365 Nitrite and sulfite
          reductases iron-sulfur/siroheme-binding site"
          /codon_start=1
          /transl_table=11
          /product="putative sulphite reductase beta subunit"
          /protein_id="CAB84608.1"
          /db_xref="GI:7380030"
          /db_xref="SPTREMBL:Q9JUD9"
          /translation="MTVOTKTKGLAMOEKPLSDNERLKTESNFTLGLTLLDKDEPLTG
          GFKGDNEQLIRFHGMYEODDDIRAEAREAKLEPLKMLRCRLPGGIIKRSOWTELD
          KFARENSHYRISRLNRQTFQHGVPKAKLQTMRLHLKLGDSIATAAMNMRVLCY
          SNPIESELHROAYEYAKKISEHLPTRTGYLDVWVGKVOSSDPELOEDEPLIGKTY
          LPKREKTVVILPPLNDVDCYGNLDLFWAISDNGOLAGFNVLGGISLMEHGNTKTY
          NISLIGFPPPEHAKAAEAVTTQRODGNSSDKMATRTTIOMGIDNFRAYERR
          MCMPEPIRPKPTKRGCRITKVGKIDGNMHLTLPESGRVDSGKOLLTGVLTIK
          IHKGFRTTAMONLIVANVPEGAKARIEQLRSGYLHIDGVSKLENNASCVSEPTCC
          LAMAEAEVRLDFITGELDKIMAKHGTSDDYIVTRITGCPNGCGRAMLEIGLVKAVG
          RYNLHIGDREGVRIPLRYKENITLPELILAEILDIGKMAERNIGEGGFPAIRGTGI
          VKPVLNAPVDFWDSKAVAIRA"
          complement(258579..260237)
          /gene="cysI"
          /note="Pfam match to entry PF01077 NIR_SIR, Nitrite and
          sulphite reductase, score 769.30, E-value 1.5e-227"
          complement(258786..258836)
          /gene="cysI"
          /note="PS00365 Nitrite and sulfite reductases
          iron-sulfur/siroheme-binding site"
          259542..259551
          /label=DUS
          /note="Core DNA uptake sequence: gccgtctgaa"
          /complement(260300..260304)
          complement(260318..262132)
          /gene="cysJ"
          complement(260318..262132)
          /gene="cysJ"
          /EC_number="1.8.1.2"
          /note="NMA1363, cysJ, probable sulphite reductase alpha
          subunit, len: 604 aa; similar to many e.g. SW:CYSJ_ECOLI
          (EMBL:M23008), cysJ, Escherichia coli sulfite reductase
          [NADPH] flavoprotein alpha-component (EC 1.8.1.2) (598
          aa), fasta scores: E(): 0, 55.7% identity in 600 aa
          overlap. Contains Pfam matches to entry PF00667
          FAD binding, FAD binding domain and to entry PF00175
          oxidored_fad, Oxidoreductase FAD/NAD-binding domain"
          /codon_start=1
          /transl_table=11
          /product="putative sulphite reductase alpha subunit"
          /protein_id="CAB84609.1"
          /db_xref="GI:7380031"
          /db_xref="SPTREMBL:Q9JUD8"
          /translation="MSEHDMONTNPPPLPMPPEITQLLSGIDAQMWLSCYAWAKAG
          NGASAGPALQOTALPTAEFPSTVLASQGNNAKSVADKADSLAAGIQVRAELD
          YKAKNINGERRLLVITSGEGEPEPAVVYLKLNKAKKAPLDKLOEAFVIGLDSY
          PNCRCACKDFDKRPEELGAKRILREVPADLDEFAAADGMDNIALKKEEAKRRARP
          APOTPPAGLOTAPDGRYCKADPPRALANOKRTTAAOSKDVNHTIEDLSGLHLT
          PGDALGYWPNDRALVREITDLGIDGATEIDAGGKTLPAVASALSHPELTQNPAY
          KGTAEPADDELDRLIADNANVLQGEVOSTPIADVLRPAKLTAEQFAGLIRPLAPRL
          YSISSSQAEVGEDEHVLTVGAVFREHEGRARAAGSGLADRLLEDGTVRVFVEENDEG
          RLPEDSRRKPIVMIGSGTVGAFFRAFYVOORAAENAGKNMLPFGNPHRPARDPIYOTEMQ
          OFAKDGFLHRYDFPAWSRDQEKIYVDKIRBQAGLQMWLODEGHNTIYVCGDAKAMAD

```

```

misc-feature      /note="Pfam match to entry PF00667 FAD-binding, FAD
                  binding domain, score 139.50, E-value 6.1e-38",
                  E-value 2.9e-38"
                  complement(260846..261433)
                  /gene="cysJ"
                  complement(260846..261433)
                  /gene="cysJ"
                  /note="Pfam match to entry PF00667 FAD-binding, FAD
                  binding domain, score 139.50, E-value 6.1e-38",
                  E-value 2.9e-38"
                  complement(260846..261433)
                  /gene="cysJ"
                  complement(260846..261433)
                  /label=DUS
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  261963..261972
                  /note="Core DNA uptake sequence: gccgtctgaa"
                  complement(262129..263415)
                  /gene="cysN"
                  complement(262129..263415)
                  /gene="cysN"
                  /EC_number="2.7.7.4"
                  /note="NMA1364, cysN, probable sulphate adenylyate
                  transferase subunit 1, len: 428 aa; similar to many e.g.
                  SW:CYSN_ECOLI (EMBL:M74586), cysN, Escherichia coli
                  sulfate adenylyate transferase subunit 1 (EC 2.7.7.4) (475
                  aa), fasta scores: E(): 0, 46.8% identity in 432 aa
                  overlap. Contains Pfam match to entry PF00009 GTP_EFTU,
                  Elongation factor Tu family, PS00301 GTP-binding
                  elongation factors signature and PS00017 ATP/GTP-binding
                  site motif A (P-loop)"
                  /codon_start=1
                  /transl_table=11
                  /product="putative sulphate adenylyate transferase subunit
                  1"
                  /protein_id="CAB84610.1"
                  /db_xref="GI:7380032"
                  /db_xref="SPTREMBL:Q9JUD7"
                  /translation="MTAQHQPPLRFRTAGSVDDGKSTLIGRLLYDSKALLSDQIKTLE
                  SGKSGIKETIDFSLITDGLEREBOGTITIDAVYRFTAKKRFITADTPGHEQYTRN
                  MWTGASTSAAVVLVDASDLQDFGAPQLLEPOTKRHSAIILRODCPHTVAVANKKDL
                  DYSEDKFNAIVEAYRRRLAEQGLKKAHVVPYSAIIGDINIVYDGMMPYKCEPLISIL
                  ETIAGDESVRADDFRPNOLYVQDDAKDPFRGYGRIRGSRVANGQVRIEPPG
                  LITAESELIITPKGEVQAQFAGSAATLRIDRDIVDSKGLFVDKNSPLAPQNHLENTLC
                  WDESRPLTARKYLLKHGTQTPAVKVGIEVSLDVRLTEQEGASLKMNDIAKVRIN
                  LQKPYTATPYAENPAAGSFIIIDEATYGTVAAGMIL"
                  complement(262204..263400)
                  /gene="cysN"
                  /note="Pfam match to entry PF00009 GTP_EFTU, Elongation
                  factor Tu family, score 279.00, E-value 6.3e-80"
                  complement(263185..263232)
                  /gene="cysN"
                  /note="PS00301 GTP-binding elongation factors signature"
                  complement(263350..263373)
                  /gene="cysN"
                  /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
                  complement(263421..263426)
                  complement(263454..264377)
                  /gene="cysD"
                  complement(263454..264377)
                  /gene="cysD"
                  /EC_number="2.7.7.4"
                  /note="NMA1365, cysD, probable sulphate adenylyate
                  transferase subunit 2, len: 307 aa; similar to many e.g.
                  SW:CYSJ_ECOLI (EMBL:M74586), cysD, Escherichia coli
                  sulphate adenylyate transferase subunit 2 (EC 2.7.7.4) (302
                  aa), fasta scores: E(): 0, 48.2% identity in 301 aa
                  overlap. Contains Pfam match to entry PF01507 PAPS_reduct,
                  Phosphoadenosine phosphosulfate reductase family"
                  /codon_start=1
                  /transl_table=11
                  /product="putative sulphate adenylyate transferase subunit
                  2"
                  /protein_id="CAB84611.1"

```

misc_feature
252814. .253527
/gene="mpl"
/note="Pfam match to entry PF01225 Mur_1lgase, Mur Ligase family, score 222.90, E-value 4.6e-63"
253161. .253170
/gene="mpl"
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
complement(253148. .253427)
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
253878. .253912
/label=DUS
/note="Stem loop containing DNA uptake sequences: gccgat gccgcctgaa agccc ttccagagc atcgc"
253883. .253892
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
complement(253898. .253907)
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
misc_feature
253945. .254061
/gene="NMA1357"
253945. .254061
/gene="NMA1357"
/note="NMA1357, len: 38 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1357"
/protein_id="CAB84603.1"
/db_xref="GI:7380025"
/db_xref="SPTREMBL:O9JUE4"
/translation="MPRIFFKRYTLAAVFSHLITRRNRHCAGLHVSIAPL"
254079. .255131
/gene="biob"
254079. .255131
/gene="biob"
/EC_number="2.8.1.6"
/note="NMA1358, biob, probable biotin synthase, len: 350 aa; similar to many e.g. SW:BI0B_ECOLI (EMBL:J04423), biob, Escherichia coli biotin synthase (EC 2.8.1.6) (346 aa), fasta scores: E(): 0, 59.1% identity in 308 aa overlap"
/codon_start=1
/transl_table=11
/product="putative biotin synthase"
/protein_id="CAB84604.1"
/db_xref="GI:7380026"
/db_xref="SPTREMBL:O9JUE3"
/translation="MTVSPVALRQETRECKKPPHPTARYMKCKQVLEAFGLPFLDIYQAA
ETHRONENPREQLQSLTSLITGCGPRECCATCPSSAHNHNTRLGKEQMMDDVEYERKK
IAKSDGNARFCGGAAMRGPRKRDVETYSAILKAVGGLGEMTCGFGLIEGEMADELKE
AGLADYNNRDLDPDRNDLIHTRHEHEDRMDTLGKVRNAGLKVCCGIVGMNENRARER
AGLISLANLDPQPEASVPIINRLVVEGEPRLDDAEDLDMTEFVARTIAVARTIMPOSYVR
LSAGSNNEPMQACPCFMAGANSIFYGDKILTTGNPEDDGRITMEKLNLYPLQFELE
GVAVAEKASGIXKYD"
255139. .255183
/note="Stem loop containing DNA uptake sequences: aaaa
gccgcctgaa accccgaaaaagcgt ttccagagcgc atttg"
255144. .255153
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
misc_feature
255169. .255178
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS
misc_feature
255169. .255178
/note="Core DNA uptake sequence: gccgcctgaa"
/label=DUS

[illegible]

```
972 CTCCTCCCATCTGATCTGTGAGAACTCTGGGGAGCTCACTGGGCG
...
2 matches found in sequence:
aaf82615 : Murine TH1 specific 200 gene coding sequence.
(from "mycobacterng.seq")
TOIG of: aaf82615 check: 4217 from: 1 to: 843

ID AAF82615 standard; cDNA; 843 BP.
XX
XX AAF82615;
AC
XX
XX 18-JUN-2001 (first entry)
DT
XX
XX Murine TH1 specific 200 gene coding sequence.
DE
XX
XX Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW anti-inflammatory; antiallergic; dermatological; antiviral;
KW antibacterial; T helper lymphocyte modulator; gene therapy;
KW TH specific gene; 200 gene; immune disorder; inflammation;
KW infection; ss.
XX
XX Mus sp.
OS
XX
XX
FH Key Location/Qualifiers
FT CDS 1..843
FT /tag= a
FT /product= "200 gene product"
PN US6204371-B1.
XX
PD 20-MAR-2001.
XX
XX 01-MAR-1996; 96US-0609583.
XX
XX 03-MAR-1995; 95US-0398633.
XX
XX 07-JUN-1995; 95US-0487748.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Levinson DA;
PI
XX
XX MPI; 2001-272703/28.
DR
XX
XX New murine or human 200 genes and their corresponding polypeptides,
PT useful for treating or diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX Claim 6; Fig 17; 109pp; English.
PS
XX
XX The present sequence is expressed at higher levels in T helper 1 (TH1)
CC cells than in T helper 2 (TH2) cells. The invention relates to an
CC isolated nucleic acid molecule, which comprises the full length murine
CC 200 gene or full length human 200 gene nucleotide sequence. The nucleic
CC acids are useful for treating or diagnosing immune disorders,
CC especially T helper lymphocyte-related disorders, e.g. inflammatory
CC diseases (e.g. Crohn's disease), multiple sclerosis, Grave's disease,
CC contact dermatitis, psoriasis, asthma and allergies, or certain viral
CC (e.g. HIV) or bacterial (e.g. tuberculosis) infections.
XX
XX Sequence 843 BP; 253 A; 185 C; 196 G; 209 T; 0 other;
SQ
AAF82615 Length: 843 March 5, 2002 14:18 Type: N Check: 4217
Found using 'seq2-3' (pappu403.key)
...
209 CTCGTAAGAAATGTGACATATCAGAAATCCAGACATACCACTAAAGCGCATCTCA
|-----|
259 264
```

```
269 ACAAGAGATGTGTCTGTGATCATTAAGAAATGTACTGTGATGA
...
514 AATGGAACAAAATAATTTCCACATGGGCTGATGAATTAAGGACTCTGGAGAAACGATCAGA
|-----|
564 569
574 ACTGCTATCCACATTGGAGTGGAGCTCTGCTGGGTGACCCCTGG
...
4 matches found in sequence:
aaf82616 : Human TH1 specific 200 gene full length cDNA.
(from "mycobacterng.seq")
TOIG of: aaf82616 check: 7978 from: 1 to: 2236

ID AAF82616 standard; cDNA; 2236 BP.
XX
XX AAF82616;
AC
XX
XX 18-JUN-2001 (first entry)
DT
XX
XX Human TH1 specific 200 gene full length cDNA.
DE
XX
XX Human; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW anti-inflammatory; antiallergic; dermatological; antiviral;
KW antibacterial; T helper lymphocyte modulator; gene therapy;
KW TH specific gene; 200 gene; immune disorder; inflammation;
KW infection; ss.
XX
XX Homo sapiens.
OS
XX
XX
FH Key Location/Qualifiers
FT CDS 42..947
FT /tag= a
FT /product= "200 gene product"
PN US6204371-B1.
XX
PD 20-MAR-2001.
XX
XX 01-MAR-1996; 96US-0609583.
XX
XX 03-MAR-1995; 95US-0398633.
XX
XX 07-JUN-1995; 95US-0487748.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Levinson DA;
PI
XX
XX MPI; 2001-272703/28.
DR
XX
XX P-PADB; AAB81518.
DR
XX
XX New murine or human 200 genes and their corresponding polypeptides,
PT useful for treating or diagnosing immune disorders, especially T helper
PT lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX Claim 1; Fig 24; 109pp; English.
PS
XX
XX The present cDNA sequence corresponds to the human 200 gene. The 200
CC gene is expressed at higher levels in T helper (TH)1 cells than in TH2
CC cells. The invention relates to an isolated nucleic acid molecule that
CC comprises the full length murine 200 gene or full length human 200
CC gene nucleotide sequence. The nucleic acids are useful for treating
CC or diagnosing immune disorders, especially T helper
CC lymphocyte-related disorders, e.g. inflammatory diseases (e.g. Crohn's
CC disease), multiple sclerosis, Grave's disease, contact dermatitis,
CC psoriasis, asthma and allergies, or certain viral (e.g. HIV) or
CC bacterial (e.g. tuberculosis) infections.
```

```

XX      SQ      Sequence 2236 BP; 586 A; 518 C; 525 G; 606 T; 1 other;
AF82616 Length: 2236 March 5, 2002 14:18 Type: N Check: 7978
Found using 'seq2-3' (pappu403.key)
...
649      TCTACATGCGAGCAGGATCTGTGCTGGGCTGCTGCTCTTATCTTGGCGCTTTAA
                                         |-----|
                                         699 704
709      TTTTCAATGGTATTCTCATAGCAAGAAGATACAGAATTAG
...
878      GTATTATTGCTATGTCAGCAGCAGCAGCACCCTCAGACACCTTTGGTGTGCGCTTGC
                                         |-----|
                                         928 933
938      AATGCATAGATCCACACCACTTATTTTGAAGCTTGTTTTC
...
1434     AGTGCATAGCGACAATCTCGGCTCAGTCAGCTCCGCCCTCGGCTTCAGCGCTTCTC
                                         |-----|
                                         1484 1489
1494     CTGGCTCAGCCTCCTGAGTGAGTGCTGGATTACAGCATGCACACCA
...
2022     TCTGTGCTGTGTGGAAGAAGCAAGCCTTCAGCATCTATATTACCGCGCTGAT
                                         |-----|
                                         2072 2077
2082     CCTTGACAGAGACTGCTCCCTAACTTAATTTCAGACGCTATA
...
-----
1 match found in sequence:
aaf82618 ; Murine TH1 specific 200 gene reverse oligo.
(from "mycobacterieng.seq")
TOIG of: aaf82618 check: 7163 from: 1 to: 39
ID      AAF82618 standard; DNA: 39 BP.
XX
AC      AAF82618;
XX
DT      18-JUN-2001 (first entry)
XX
DE      Murine TH1 specific 200 gene reverse oligo.
XX
KW      Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW      anti-inflammatory; antiallergic; dermatological; antiviral;
KW      antibacterial; T helper lymphocyte modulator; gene therapy;
KW      TH specific gene; 200 gene; immune disorder; inflammation;
KW      infection; PCR primer; ss.
XX
OS      Mus sp.
XX
PN      US6204371-B1.
XX
PD      20-MAR-2001.
XX
PF      01-MAR-1996; 960S-0609583.
XX
PR      03-MAR-1995; 950S-0398633.
XX      07-JUN-1995; 950S-0487748.
XX      (MILL-) MILLENNIUM PHARM INC.
PA
PS

```

```

PI      Levinson DA;
XX
DR      WPI: 2001-272703/28.
XX
PT      New murine or human 200 genes and their corresponding polypeptides,
PT      useful for treating or diagnosing immune disorders, especially T helper
PT      lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX      Example: Column 89; 109pp; English.
XX
XX      The present sequence was used to amplify the signal sequence and
XX      extracellular domain of the murine 200 gene. The invention relates to an
XX      isolated nucleic acid molecule, which comprises the full length murine
XX      200 gene or full length human 200 gene nucleotide sequence. The nucleic
XX      acids are useful for treating or diagnosing immune disorders,
XX      especially T helper lymphocyte-related disorders, e.g. inflammatory
XX      diseases (e.g. Crohn's disease), multiple sclerosis, Grave's disease,
XX      contact dermatitis, psoriasis, asthma and allergies, or certain viral
XX      (e.g. HIV) or bacterial (e.g. tuberculosis) infections.
XX
SQ      Sequence 39 BP; 7 A; 10 C; 7 G; 15 T; 0 other;
AF82618 Length: 39 March 5, 2002 14:18 Type: N Check: 7163
Found using 'seq2-3' (pappu403.key)
...
1      TTAATTGATCCCGATTGATGCTTTCACAGATC
                                         |-----|
                                         23 28
-----
1 match found in sequence:
aaf82619 ; Human TH1 specific 200 gene forward oligo.
(from "mycobacterieng.seq")
TOIG of: aaf82619 check: 7788 from: 1 to: 32
ID      AAF82619 standard; DNA: 32 BP.
XX
AC      AAF82619;
XX
DT      18-JUN-2001 (first entry)
XX
DE      Human TH1 specific 200 gene forward oligo.
XX
KW      Human; T helper cell; TH cell; TH1; TH2; immunomodulator;
KW      anti-inflammatory; antiallergic; dermatological; antiviral;
KW      antibacterial; T helper lymphocyte modulator; gene therapy;
KW      TH specific gene; 200 gene; immune disorder; inflammation;
KW      infection; PCR primer; ss.
XX
OS      Homo sapiens.
XX
PN      US6204371-B1.
XX
PD      20-MAR-2001.
XX
PF      01-MAR-1996; 960S-0609583.
XX
PR      03-MAR-1995; 950S-0398633.
XX      07-JUN-1995; 950S-0487748.
XX      (MILL-) MILLENNIUM PHARM INC.
PA
PS      Levinson DA;
XX
XX      WPI: 2001-272703/28.
XX
XX      New murine or human 200 genes and their corresponding polypeptides,
XX      useful for treating or diagnosing immune disorders, especially T helper
XX      lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX      Example: Column 89; 109pp; English.
XX

```

```

XX CC The present sequence was used to amplify the signal sequence and
CC CC extracellular domain of the human 200 gene. The human 200
CC CC gene is expressed at higher levels in T helper (TH)1 cells than in TH2
CC CC cells. The invention relates to an isolated nucleic acid molecule that
CC CC comprises the full length murine 200 gene or full length human 200
CC CC gene nucleotide sequence. The nucleic acids are useful for treating
CC CC or diagnosing immune disorders, especially T helper
CC CC lymphocyte-related disorders, e.g. inflammatory diseases (e.g. Crohn's
CC CC disease), multiple sclerosis, Grave's disease, contact dermatitis,
CC CC psoriasis, asthma and allergies, or certain viral (e.g. HIV) or
CC CC bacterial (e.g. tuberculosis) infections.
XX SQ Sequence 32 BP; 9 A; 7 C; 7 G; 9 T; 0 other;
AAAF82619 length: 32 March 5, 2002 14:18 Type: N Check: 7788
Found using 'seq2-3' (pappu403.key)

1 -----|
AAATTTATCTCGAGCGCTAACAGAGGTGTC
14 19

1 match found in sequence:
aaf82621; Murine TH2-enriched 103 gene 5' oligonucleotide.
(from "mycobacterng.seq")
TOIG of: aaf82621 check: 5916 from: 1 to: 31

ID AAF82621 standard; DNA; 31 BP.
XX AAF82621;
AC AAF82621;
XX
XX 18-JUN-2001 (first entry)
DT
XX Murine TH2-enriched 103 gene 5' oligonucleotide.
DE
XX Murine; T helper cell; TH cell; TH1; TH2; immunomodulator;
XX anti-inflammatory; antiallergic; dermatological; antiviral;
XX antibacterial; T helper lymphocyte modulator; gene therapy;
XX TH2 enriched; 200 gene; immune disorder; inflammation;
XX infection; 103 gene; PCR primer; ss.
XX
XX Mus sp.
OS
XX US6204371-BL.
PN
XX 20-MAR-2001.
PD
XX
XX 01-MAR-1996; 96US-0609583.
PF
XX
XX 03-MAR-1995; 95US-0398633.
PR
XX 07-JUN-1995; 95US-0487748.
PT
XX (MILL-) MILENNIUM PHARM INC.
PA
XX
XX Levinson DA;
PI
XX WPI; 2001-272703/28.
DR
XX
XX New murine or human 200 genes and their corresponding polypeptides,
XX useful for treating or diagnosing immune disorders, especially T helper
XX lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX Example; Column 90; 109pp; English.
PS
XX The present sequence was used to amplify a DNA fragment encoding the
XX extracellular domain of the murine 103 gene, a T helper 2 (TH2) enriched
XX gene. The invention relates to an isolated nucleic acid molecule, which
XX comprises the full length murine 200 gene or full length human 200 gene
XX nucleotide sequence. The nucleic acids are useful for treating or
XX diagnosing immune disorders, especially T helper lymphocyte-related
XX disorders, e.g. inflammatory diseases (e.g. Crohn's disease), multiple
XX disorders, e.g. inflammatory diseases (e.g. Crohn's disease), multiple

```

```

CC CC sclerosis, Grave's disease, contact dermatitis, psoriasis, asthma and
CC CC allergies, or certain viral (e.g. HIV) or bacterial (e.g. tuberculosis)
CC CC infections.
XX SQ Sequence 31 BP; 5 A; 8 C; 12 G; 6 T; 0 other;
AAAF82621 length: 31 March 5, 2002 14:18 Type: N Check: 5916
Found using 'seq2-3' (pappu403.key)

1 -----|
CCGCGGTACCAATTCGTCGGGGTGC
17 22

1 match found in sequence:
aaf82627; Murine 103 gene transgenic construct pCD2-103L-GH 5' oligo.
(from "mycobacterng.seq")
TOIG of: aaf82627 check: 6990 from: 1 to: 21

ID AAF82627 standard; DNA; 21 BP.
XX AAF82627;
AC AAF82627;
XX
XX 18-JUN-2001 (first entry)
DT
XX Murine 103 gene transgenic construct pCD2-103L-GH 5' oligo.
DE
XX
XX T helper cell; TH cell; TH1; TH2; immunomodulator;
XX anti-inflammatory; antiallergic; dermatological; antiviral;
XX antibacterial; T helper lymphocyte modulator; gene therapy;
XX TH2 enriched; 200 gene; immune disorder; inflammation;
XX infection; 103 gene; PCR primer; ss.
XX
XX Synthetic.
OS
XX US6204371-BL.
PN
XX 20-MAR-2001.
PD
XX
XX 01-MAR-1996; 96US-0609583.
PF
XX
XX 03-MAR-1995; 95US-0398633.
PR
XX 07-JUN-1995; 95US-0487748.
PT
XX (MILL-) MILENNIUM PHARM INC.
PA
XX
XX Levinson DA;
PI
XX WPI; 2001-272703/28.
DR
XX
XX New murine or human 200 genes and their corresponding polypeptides,
XX useful for treating or diagnosing immune disorders, especially T helper
XX lymphocyte-related disorders, e.g. inflammatory diseases or allergies
XX
XX Example; Column 94; 109pp; English.
PS
XX The present sequence was used to generate a 32P-radioabelled PCR
XX fragment of the pCD2-103L-GH construct, which contains the coding
XX sequence of the murine T helper 2 (TH2) enriched 103 gene.
XX The invention relates to an isolated nucleic acid molecule, which
XX comprises the full length murine 200 gene or full length human 200 gene
XX nucleotide sequence. The nucleic acids are useful for treating or
XX diagnosing immune disorders, especially T helper lymphocyte-related
XX disorders, e.g. inflammatory diseases (e.g. Crohn's disease), multiple
XX sclerosis, Grave's disease, contact dermatitis, psoriasis, asthma and
XX allergies, or certain viral (e.g. HIV) or bacterial (e.g. tuberculosis)
XX infections.
XX SQ Sequence 21 BP; 3 A; 4 C; 8 G; 6 T; 0 other;
AAAF82627 length: 21 March 5, 2002 14:18 Type: N Check: 6990
Found using 'seq2-3' (pappu403.key)

```


CC patient having a Type 1 interferon treatable disease. They are useful
 CC for treating autoimmune, mycobacterial, neurodegenerative, parasitic or
 CC viral disease, arthritis, diabetes, lupus multiple sclerosis, leprosy,
 CC tuberculosis, encephalitis, malaria, cervical cancer, genital herpes,
 CC hepatitis B or C, human immunodeficiency virus (HIV), human papilloma
 CC virus (HPV), herpes simplex virus (HSV)-1 or 2, or neoplastic disease
 CC such as multiple myeloma, hairy cell leukemia, chronic myelogenous
 CC leukemia, low grade lymphoma, cutaneous T-cell lymphoma, carcinoma
 CC tumours, cervical cancer, sarcomas including Kaposi's sarcoma, kidney
 CC tumours, carcinomas including renal cell carcinoma, hepatic cellular
 CC carcinoma etc.

XX Sequence 1014 BP; 305 A; 222 C; 240 G; 247 T; 0 other;

AA90293 Length: 1014 March 5, 2002 14:18 Type: N Check: 6359 ..
 Found using 'seq2-3' (pappu403.key)

79 GAGCAGACATTTCAGAACTATCCAGAGGCAAAACCCCGGCCACATGAGAGCTGANG
 129 134

139 TTGGATGGCAACCTTCAGCTAGACTGCTGCTCAAGGGGTGAAGC

1 match found in sequence:

aah73439 ; Immunomodulatory nucleic acid.

(from "mycobacterieng.seq")

TOIG of: aah73439 check: 8143 from: 1 to: 22

ID AAH73439 standard; DNA; 22 BP.

AC AAH73439;

DT 01-OCT-2001 (first entry)

DE Immunomodulatory nucleic acid.

KW G3PDH gene; immunomodulatory oligonucleotide; infection; mycobacterium;

KW intracellular pathogen; anti-pathogenic; ss.

OS Unidentified.

PN WO200155341-A2.

PD 02-AUG-2001.

PF 30-JAN-2001; 2001WO-US03029.

PR 31-JAN-2000; 2000US-0179353.

PA (REGC) UNITV CALIFORNIA.

PI Raz E, Kornbluth R, Catanzaro A, Hayashi T, Carson DA;

DR WPI; 2001-483234/52.

PT treating infection of intracellular pathogen e.g., Mycobacterium, in a
 PT subject, involves administering immunomodulatory nucleic acid molecule
 PT to inhibit intracellular replication of intracellular pathogen

PS Examples; Page 26; 54pp; English.

CC The present invention describes a method of treating an infection caused
 CC by an intracellular pathogen, involving administering to the patient an
 CC immunomodulatory nucleic acid and an anti-pathogenic agent. This is
 CC particularly useful in the treatment of mycobacterial infections. The
 CC present sequence is an immunomodulatory nucleic acid described in the
 CC exemplification of the invention.

SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
 AAH73439 Length: 22 March 5, 2002 14:18 Type: N Check: 8143 ..
 Found using 'seq2-3' (pappu403.key)

1 TGACTGTGAACCTTCAGATGA
 9 14

67 matches found in sequence:

aaq81456 ; Orotidine-5'-monophosphate-decarboxylase gene.

(from "mycobacterieng.seq")

TOIG of: aaq81456 check: 9478 from: 1 to: 4393

ID AAQ81456 standard; DNA; 4393 BP.

AC AAQ81456;

DT 19-AUG-1995 (first entry)

DE Orotidine-5'-monophosphate-decarboxylase gene.

KW Orotidine-5'-monophosphate-decarboxylase gene; recombinant vaccine;

KW attenuation; BCG; selectable marker; ura gene inactivation;

KW homologous recombination; carbamoylphosphate-synthase large subunit;

KW Mycobacterium tuberculosis; Mycobacterium leprae; antigen; ds.

OS Mycobacterium bovis (BCG).

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Slow-growing mycobacteria transformed with heterologous DNA -
 useful as vaccines in which the mycobacteria stimulate long-term
 memory or immunity

Claim 36; Page 37-39; 59pp; English.

CC The sequence contains a Mycobacterium bovis BCG ura gene,
 CC encoding orotidine-5'-monophosphate-decarboxylase (OMD), and
 CC flanking sequences useful in homologous recombination, including
 CC an upstream open reading frame encoding the C-terminus of a
 CC protein (AAR76075) homologous to a carbamoylphosphate-synthase
 CC large subunit, and a downstream open reading frame encoding an
 CC N-terminal segment (AAR76076) homologous to Mycobacterium
 CC tuberculosis and Mycobacterium leprae antigens. The gene may be
 CC inactivated or deleted in BCG to produce a slow-growing

CC attenuated strain, which is useful as a vehicle for recombinant
CC vaccine production. The gene may also be used as a selectable
CC marker.
xx
SO Sequence 4393 BP; 684 A; 1446 C; 1508 G; 755 T; 0 other;

AA081456 Length: 4393 March 5, 2002 14:18 Type: N Check: 9478
Found using 'seq2-3' (pappu03.key)

```
...
187 ATGATGTCACCGACTTCGACACCGCGGACAGAGTTGTAATTGACCCGTTGACGTTGAGG      |-----|
247 GACGCTTTGGAGGTTCTACCAACCGCAAAATGGAATCCGTTAGCGGTGCGCCGGAGTGGCC      |-----|
307 GCGGTATCGTTCAGACTGGGGGCGCAAGACCCCGCTCGGCTGGCGACCGGCTCCGCCGACG      |-----|
367 CCGGTCCTCCGCTCGTGGGCAACCAACGGAGSCATTCGACTGGCCGAGGATGGCCCGTT      |-----|
427 CGGCGACTCTGCTGAGCGAGACTGCCGCGCCAAAGTAAGGACCGACCGCAACTTTTCGCC      |-----|
487 CAGGCCCGCGGATCGCCGAGAGATCGGCTATCCGGTGTGTGCGGCCCTCGTATGTG      |-----|
547 CTCGGTGTGCGCGCATGAGANTGTGTATGACGAGAAACGTTGTCAGAGGCTTACATCAC      |-----|
607 CGCGCCACTCAGCTATCCCGCGAACCACCGGTGCTCGTGACCGCTTCCTCGAGAGACGG      |-----|
667 GTGAGATTCGACGCTGTCTGTGTGATGGCGCGGAGTCTATATCGGGGGAATCATG      |-----|
727 GAGCATATCAGAGAGCGCGCATCCACTCGGTGACTCGGCTGTGCGGTTCACACGGTTC      |-----|
787 ACGTTGGGGCGGAGCGCATTCGAGAAAGGTGGCTAAGGCCACTGAAGCCATTGGCGCATGG      |-----|
847 ATCGGCGT 799 804
...
950 CCAAGGCCACAGCGGTGCACTGCGCAAGGCATGCGCCCGGATCATGTTGGGCGCACCA      |-----|
1010 TTGCCCAAGTGGCGCGCAAGGCTTGCTGGCGGTACACGGGGAGTGGCGCCACGCGCGC      |-----|
1070 GAAAGCGCCCGCATCGCGGTACACCAAGCCGTTTGGCGTTTCAACCGGTTCCGGCGCGC      |-----|
1195 GACCGGAGCTTGGGACAGCGGTTTCGCCAAGAGCCAGACCGCGCCTAAGGGTGGTGGCC      |-----|

1255 GCCCAGGGCACAGTGTGTGTGGTGGCCAACCGGGACAAAGCGGTCGCTGTGTTCCG      |-----|
1315 GTCAACGATTGGCCCGCCACCTGGGTTTTCGCGTCTTCGCCACCGAAGCACACCGAGANT      |-----|
1375 TCGCGCCGACAGGTTATTCCTCCGACAGACTCCGCAAAATTTGAGCGCGGCGACCCG      |-----|
1435 GCCGCCCCCAATGTTCGGCGGT 1398 1403 1401 1406
1582 CGGTGCAAGGGGCATCCGCCCGCTGCAGGGGATGAGGCGCGGATCCGCGGCGACATCG      |-----|
1642 GGGTGCCTCCCTCGAGAGAGCTGCACCGGGTGATCGGGGGGCTCGAGCGGTGACCGGTT      |-----|
1702 CGGTCTCCGTTGGCCGAGGCAAAAGGCACGCCGC 1680 1685
...
1742 TTGTGTGGGCATTCGATCCGATCCGAGCTGTGCGGGGCTGGATCGGCGACACG      |-----|
1802 GCCACAGGCTGGCGCGCTTCTGCGACATCTGCCGTACGGGCTTCGCTGATTTCGCGGTG      |-----|
1862 GTCAACCGCAGGTGGGCTTTTGTGAGTCAATACGGGCTGCGGATTCGCGGTGTGAG      |-----|
1922 CGCACATTCGCGAATGCGGGCGCGCAGACGTGCTGTGTTGGCGGACCGACCGCGC      |-----|
1982 GACATTGGGGGAGCACATTCGGCTATGCGAGCGCTGGTGGCGGACTCGCCCTGGCC      |-----|
2042 GCCAGCGCGTGAACCGCGCTCGCCCTATTGGGTTTCGTTGCTGGCGCGCTCTAGA      |-----|
2223 TGGGGGCGCCAAAGAGGCGGAGAGACCGGGCGCGGATCCATTCAGCGGTGGTGGCG      |-----|
2283 CAACGGGCGCCAGGCGCCCGGATTCAGCGCTTTCACCGGGCGGTGCTGTGGCCCGCG      |-----|
2343 TGGGGGTGACAGGCGGCGCGCCCGGAGGCGCTTGGGCGGGGCGCATGAGCC      |-----|
2403 AGCTGTTCGCCGCGGTGGCGC 2358 2363 2373
...

```

```

2442  GCCCGCGCTGCCCGAATTGGCGCGCGGCGAGCAGATGCGCATGCCGCTGATC      2492 2497
2502  TGGCTGCCGTGATAGCGGGTGCCCTGCCACCGCGCCGCTAAATCCACACAGCATGGGGTG
2562  TGAGCCCGACGCGCTGCTGTGACCAAACTCACCGCCCTGGGCGCGCTGCACCTGTGTTAAC      2603 2608
2622  CTCTCGTCAAAATGATATTTCATATTCAATAGTGGCGCTAGTGTCCGTTGAATCCCGGT      2654 2659
2682  TGAACCCCAACAGATGAGTCTGTGTCTCGTAGTGTGCGAGTGTGTTCCCAAGCCTGCG      2712 2717 2722 2727
2742  AGCGCGCCAGCGGCTGCCATCGAAGCAGTGAACCGCTCGGCTGGCGCGCGCGCGCGGCG      2743 2748
2802  GGC      2749 2754
...
2807  CGTTTATCGCGGCGTATCCCGCTGGGTCCGACTCGGTTTGGTGTGACACGCGTTG      2857 2862
2867  AGTTCAGCGTTTCACGGTAGTCAGCATGTGGCAATGGCCGCTCAGGGGTTGAGGAGCTCG      2872 2877
2927  G
2940  CGGGGTGGCCGGAATCGGGTGCCAGTTATGCCGCTAGAGATGCGCTGGCGCGGCGTGTAT      2990 2995
3000  CTCAGGGGTGGGCTATGACCGAGCCGTGATAGCTTCCCTCCCGAGGTGCTACTCGCGCA      3055
3060  TCGTGAACATPACGGTGGCGGCTTGGGCGCATGTTGATCTCCGCCACGACAGATGGGAGC      3077 3082
3120  TCAGCGCCCAATACGAGAGAGCGCATCCGAGGTGAGAAATTGTGGGGGTGTGGCCT      3122 3127
3180  CCGAGGATGGCAGGGGCAAGCCCGAGGCGCTTAGTGCAGCTACATGCCGTTTCTGG      3208 3213 3220
3240  CGTGGCTGATCAAGCCAGCGGCTGCTGGAATGGCGCCCGCACACACGCGCTCA      3257 3262
3300  TCGAGGCGCTACAC
...
3332  CTTACTCAGGTCGAACTGGCGCCCAACCAATCAAGCTCGGCTGTGTGTAAGGACCAAT      3382 3387
3392  TTCTTTGGCATCAACACCATTCCTCATTTGCGATCAGTAGGCGCGAGTACGTGAGATGTG

```

```

3452  GTTCGGGCGCCCAACACAGATGGGACCTATTCAACAGTCTCCAGATGGGCGTCCCG      3472 3477
3512  ATGCCGCACACAGACCCCGCGCTGATCTGAAATCGATG
...
3618  ATGGCGGTCAAGCCAGATGATGATTACTTTTTCGCCAATTCCTGCGGTCAAGC      3668 3673
3678  CGGGCGCATTTGTTGGGACCCCGTCAAGCAGCAACCTCAAGGACTGCTACAGCATT      3731 3736
3738  ACGTATACCCGGTACAGCATCTGTGTGCTGAGGCGCTCGAGTT
...
3810  GCGAATGTTGTTACCAATCCAGTGGGCTTTTCACTTCTCTCTACGTCTGTG      3860 3865
3870  TGGATTGGCGAGCGACATAGCCAGATCCCTACTGTGGGCGCACTACCCGAGTTG      3896 3901
3930  TGTGCGCTGCCCTACAGCGCATGCGCCACCTGGAGCAATPACTGTTTGGCGGGC      3947 3952
3990  TATCCGCGCTGAGCGCATTCGCTCTGTGATACCCGCGTGTACCGAGCTGACAC      4001 4006
4050  CCGTCGCGGCGCGCGCCCTATGTTGGCGGTCCGCGGGTGGCCCTGCAGTCCCGC      4079 4084
4110  CGGGCATGTCTCCCGCTCAGACACCCGACCGGCGGAGCGCGCCACCGCAGCG      4154 4159
4170  GCCCGAGCGCGCGCGACTGTGTTTCGAGAGGCTTCCGCCCTACTGTGTGGGCTG      4174 4179 4188
4230  GGCCGAGGAATAGGTTGCGGCTCGGACAGTGGGCCACCGCAAGGCCCGGCGCTCCGAT      4280 4285
4290  TCCGCTGACGCGAGTGGCGGCGCCAGGCTTGGCGGCTGGCGAGG
...
37 matches found in sequence:
aas02609 ; M. tuberculosis heat shock protein 70 (HSP70) DNA sequence.
(from "mycobactering.seq")
TOIG of: aas02609 check: 3431 from: 1 to: 1878
ID AAS02609 standard; DNA: 1878 BP.
XX
XX AAS02609;
XX AC
XX DT 29-AUG-2001 (first entry)
XX DE M. tuberculosis heat shock protein 70 (HSP70) DNA sequence.
XX KM Mouse; granulocyte-macrophage-colony stimulating factor; GM-CSF;

```



```

1220 AACCGTGGTGACATCCAGGCTATCATCAGGGAGACCGTGAATCGCCCGCACACAACT
      |-----|
      1261 1266
1280 TGCTGGGCTCCTCGAGCTGACCGGCATCCCGCGCGCGGGGATTCGCGAGATCG
      |-----|
      1314 1319
1340 AGGTCACTTTGACATCGACGCCCAAGGCATTGTGCACGTACCGCCAGACAAAGGCA
      |-----|
      1357 1362
1400 CCGGCAAGAGAGA
...
1424 TCCAGGAAGGCTCGGGCCTGTCCAGGAAGACATTGACCGCATGATCAAGTCGCGGAG
      |-----|
      1474 1479
1484 CGCACGCCGAGAGAGATCGACGCGTGCAGAGAGCGCGAGTGTCTAATCAAGCCGAGA
      |-----|
      1505 1510
1544 CATTGGTCTACGACAGC
...
1645 GCGGAAGCGAAGCGCGCACTTGGCGGATGGAATATTTCGCCCATCAAGTCGGCGAGGAG
      |-----|
      1695 1700
1705 AAGCTGGGCCAGAGACTCCAGAGCTCTGSGGGCAGTCTACGAAGCAAGCTCAGGCTCGC
      |-----|
      1737 1742
1765 TCACAGGCCACTGGCGGCTGCCACCCCGCGCGAGCGGGGGTGTCCACCCCGGCTCG
      |-----|
      1777 1782
1825 GCTGATGACGTTGTGACGCGGAGGTGCTGCACGACGCGCGGAGGCCCAAGTGA
      |-----|
      1831 1836
      1855 1860

-----
38 matches found in sequence:
aas02618 ; Chimeric immunogenic protein E7-HSP70, DNA sequence.
(from "mycobacterng.seq")
ToIG of: aas02618 check: 6258 from: 1 to: 2172

ID AAS02618 standard; DNA: 2172 BP.
XX AAS02618;
AC
XX
XX 29-AUG-2001 (first entry)
DT
XX
XX Chimeric immunogenic protein E7-HSP70, DNA sequence.
DE
XX
XX Mouse; granulocyte-macrophage-colony stimulating factor; GM-CSF;
KW chimeric; heat shock protein; HSP; Flt-3 ligand; FL; exotoxin A;
KW dlt; antigenic; immunogenic; cytotoxic T cell response; tumour;
KW vaccine; immunotherapy; ds.
XX
OS Chimeric - Mycobacterium tuberculosis.
OS Chimeric - Human papilloma virus.
XX
FH Key
FT 1..2172
FT /tag= a
FT /product= "E7-HSP70 chimeric protein"
XX

```

```

PN WO200129233-A2.
XX
XX 26-APR-2001.
PD
XX 20-OCT-2000; 2000MO-US41422.
XX
XX 20-OCT-1999; 99US-0421608.
PR 09-FEB-2000; 2000US-0501097.
XX
XX (UWYO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
XX Wu T, Hung C;
PI
XX
XX WPI: 2001-290921/30.
DR
XX P-PSDB: AAD01720.
XX
XX New chimeric polypeptide, useful as anti-tumour vaccines, comprises
PT carboxy terminal fragment of heat shock protein, Flt-3 ligand or
PT cytoplasmic translocation domain of Pseudomonas exotoxin A and
PT antigenic polypeptide -
XX
XX Example 1; Fig 13; 110pp; English.
PS
XX
XX The sequence represents the coding sequence of a chimeric polypeptide
CC comprising: (a) a first polypeptide domain containing a carboxy terminal
CC fragment of a heat shock protein-70 (HSP70); and (b) a second polypeptide
CC domain containing human papilloma virus-16 (HPV-16) E7 antigenic
CC polypeptide. A composition comprising the chimeric polypeptide is useful
CC for inducing an immune response such as a cytotoxic T cell response. The
CC nucleic acid or vector encoding the chimeric polypeptide present in the
CC composition is administered as naked DNA by gene gun or equivalent, or by
CC liposomal formulation. These are thus useful for vaccinating a mammal
CC against infection by inducing an immune response to a pathogen.
CC Preferably they are useful for vaccinating a mammal against a tumour
CC antigen. The compositions and methods are useful for stimulating or
CC enhancing the immunogenicity of a selected antigen or stimulating or
CC enhancing a cellular immune response specific for that antigen. The
CC chimeric nucleic acid molecules and vaccination methods, yield potent
CC antigen-specific immunotherapy. The polynucleotides and DNA vaccines can
CC induce a cellular immune response that is at least 40 fold more potent
CC than conventional DNA vaccines. The vaccines are safe and useful for
CC administration to domesticated or agricultural animals, as well as
CC humans, and have low immunogenicity.
XX
XX Sequence 2172 BP; 482 A; 626 C; 692 G; 369 T; 3 other.
SQ

AAS02618 Length: 2172 March 5, 2002 14:18 Type: N Check: 6258 ..
Round using 'seq2-3' (pappu403.key)

...
287 AAGATTCATGGCTGCTGCGGTGCGGATTCGACCTCGGGAGCACCAACTCCGCTCGCG
      |-----|
      337 342
347 TTCTGGAAGTGGCGACCCGGGTGCTGTCGTCGCAACTCGAGGGCTCCAGAGCACCCCGT
      |-----|
      358 363 367 370 372
      373 375
      378
407 CAATTGTGCGCTTGCCTCCGCAACGGTGAAGTGTGCTGCGGCCAGCCGCCCAAGAACAGG
      |-----|
      475 480
      503 508 517 522
467 CAGTGACCAAGCTGATCGACACCGGTGCTGCTGCAAGGACACATGGGACGACTGGT
      |-----|
      527
      568 573
CCATGAGATTGACGCGCAAGAAATATACACCGCGCCGAGATGACGCCCGGTATTCTGATGA

```

```

587      AGCTGAGGCGGACCGCCGCTACCTCGGTGAGACATTAACGACGGGTTATTCAGCA
      598 603
647      CGCCCGCCTACTTTCACGACGCCAGCGCTGAGCCACCAAGACGCCGCGATCGCG
      650 661 666 671 676 688 693 700 705
707      GCCTCAACGTGCTGGGATTCGTCAACGAGACCGACCGCGCGCTGGCTACGGCTCG
      724 729
767      ACAAGGGCGAGAGAGGACGACGAAATCTGTCTTCGACTTGGGTGTGACACTTTGAGG
      823
827      TTTCCCTCTGTGAGATCGGCGGAGGGTGTGTTGAGGTCGCTGCACCTTCGGGTGACAAC
      828
887      ACGTGGGGGGGAGCGACTGCGGACCGCGGCTGCTGCTGGTGACAACTTCAAG
      895 900 916 921
      898 903
947      GCACGAGCGGCGATCGATCTGACCAAGGACAAAGATGGCGATGAGGGCTGCGGGAGCGG
      981 986
1007      CCGAGAGGCAAGATCGAGCTGAGTTGAGTTCACCTCCACCTCGATCAACCTGCCCTACA
      1075 1080
1067      TCACCGTGCAGCGCGACAAGAACCGGTTGTCTTAGACGAGAGCTGACCGCGCGGAGCT
      1075 1080
1127      TCCTCA
...
1133      GGATCAGCTCAGAGACCTGCTGAGACCGCACTCCGAAAGCCGTTCCAGTGGCTGATCGCTGACA
      1183 1188
1193      CCGGCATTTTCGGTGTGAGATCGATCACTTGTGCTCGTGGTGG
...
1256      CGGTGACCGGATTTGTCGAAGAGACTCACCGGCGGCAAGAAACCAAGGCGCGTCAACC
      1306 1311
1316      CCGATGAGAGTTGTGGCGGTGGGAGCCGCTTGTCAAGCCGCGGCTCTCAAGGCGCAGGTGA
      1354 1359
1376      AAGACGTTTGTGCTGCTGATGTTAACCCCGGCTGAGCGTGGTATGAGACCAAGGCGCG
      1378 1383
...
1454      AGCGCAACACCAAGATCCCGCAAGAGGGTGGAGACTTTCACGACCGCGCGACAGCAAC
      1504 1509
1514      AACCGTGGGTGACAGATCCAGGCTTATCAGGGGAGCGGTGAGATCGCGCGACCAACAAGT
      1555 1560

```

```

1574      TGCTGGGTCCTTGAGCTGACCGGACCTCCCGCGCGCGGCGGATTCGCGACAGATCG
      1608 1613
1634      AGGTCACTTTCGACATCGACGCCCAAGCGCATTTGTGACGCTACCGCCAAAGACAGGCGA
      1651 1656
1694      CCGGCNAGGAGAA
...
1718      TCCAGAGAGCTCGGGCTGTCCAAAGAGACATTGACCGCATGATCAAGTCAAGCGGAG
      1768 1773
1778      CGCAGCGCGAGAGAGATCGCAAGCGTTCGGAGAGAGCGCATGTTCTTAACAACCGAGA
      1799 1804
1838      CATGTGCTTACCAGACG
...
1939      GCGAAGCGAAGCGGCGCACTTGGCGGAGTCGATATTTCGCCCATCAAGTCGGCGATGAG
      1989 1994
1999      AAGCTGGGCGCAGAGATCGCAGGCTGTGGGCAAGGATTTACGAAGCAAGCTCAGGCTGCG
      2031 2036
2059      TCACAGGCCACTGGCGCTGCCACCCCGCGGCGAGCGGCGGTGCMCACCCCGGCTCG
      2071 2076
2119      GCTGATGACGTTGTGACGCGGAGGTGTGTCGACGACGGCGCGGAGGCCAAGTGA
      2125 2130 2149 2154
-----
16 matches found in sequence:
aas02619 : Chimeric, immunogenic protein GM-ETA(dII)-E7, DNA sequence.
(from "mycobacterieng_seg")
ToIG of: aas02619 check: 7771 from: 1 to: 1263
ID      AAS02619 standard; DNA; 1263 BP.
XX
AC      AAS02619;
XX
DT      29-AUG-2001 (first entry)
XX
DE      Chimeric immunogenic protein GM-ETA(dII)-E7, DNA sequence.
XX
KW      Mouse: granulocyte-macrophage-colony stimulating factor; GM-CSF;
KW      chimeric; heat shock protein; HSP; Flt-3 ligand; FL; exotoxin A;
KW      ETA dII; antigenic; immunogenic; cytotoxic T cell response; tumour;
KW      vaccine; immunotherapy; ds.
XX
OS      Chimeric - Mycobacterium tuberculosis.
OS      Chimeric - Human papilloma virus.
OS      Chimeric - Pseudomonas sp.
XX
FH      Key
FH      1. .1263
FH      CDS
FH      Location/Qualifiers
FT      /*tag= a
FT      /product= "GM-ETA(dII)-E7 chimeric protein"
FT      /*tag= b
FT      /note= "GM-CSF domain"
FT      430..942
FT      /*tag= c
FT      /note= "ETA dII domain"
FT      949..1263
FT      misc_feature

```



```

XX 18-OCT-2000; 2000WO-US28827.
PF
XX
XX 20-OCT-1999; 99US-0160542.
PR
XX
XX (GETH ) GENENTECH INC.
PA
XX De Sauvage FJ, Grewal I, Gurney AL;
PI
XX WPI: 2001-308474/32.
DR
XX P-PSDB: AAU01854.
XX
XX Modulating T-cell differentiation and cytokine release profiles into
PT Th1 and Th2 subtypes, for treating immune-related diseases in mammals,
PT by administering modulator of type I cytokine receptor (TCCR)
XX
PS Disclosure; Page 123-125; 126pp; English.
XX
XX The sequence encodes mouse type I cytokine receptor, mTCCR. The
CC invention relates to methods of modulating the differentiation of
CC T-cells into the Th2 subtype instead of the Th1 subtype, by
CC administering a modulator of TCCR (e.g. an antagonist) to enhance,
CC stimulate or potentiate T-cell differentiation, or using TCCR
CC polypeptide or its agonists to prevent, inhibit or attenuate T-cell
CC differentiation. Th1 mediated disease in mammal can be treated by
CC administering a TCCR antagonist and Th2 diseases by administering a TCCR
CC agonist. Th1-mediated diseases include allograft rejection and autoimmune
CC inflammatory diseases, such as allergic encephalomyelitis, multiple
CC sclerosis, insulin-dependent diabetes mellitus, autoimmune uveoretinitis,
CC inflammatory bowel disease or autoimmune thyroid disease. Th2-mediated
CC diseases include infectious diseases, such as Leishmania major,
CC Mycobacterium leprae, Candida albicans, Toxoplasma gondii, respiratory
CC syncytial virus and human immunodeficiency virus (HIV) and allergic
CC disorders, such as asthma, allergic rhinitis, dermatitis and vernal
CC conjunctivitis.
XX
XX Sequence 2005 BP; 425 A; 570 C; 577 G; 433 T; 0 other:
SQ
AAS03263 Length: 2005 March 5, 2002 14:17 Type: N Check: 2233 ..
Found using 'seq2-3' (pappu403.key)
...
18 GCCATGAACCGGCTCCGGGTTCACAGCCTCAGCCGTTGGAGCTTCTGCTGCTGCTGATG
|-----|
68 73 77
78 TCGCTGCTGCTCGGGACGCGGCCCCACGAGTCAGGCCCATGCAAGTGTACAGCGTC
|-----|
82 132 137
138 GGTCCCTGGGAATCTGAACTGCTCCTGGGAACCTTTGGCGAGCTGCATCCACT
|-----|
177 182
198 GTGCTGATCACACAGTACAGAAATACCATCCAA
...
408 TCTCAAGTGATATTCTGAGAGAACACCTCGAGGCCACTGTGACAGTGGCGCGGCC
|-----|
458 463
468 GTGTGGCCACCGCAGAAAGCTCTCACTGTGAGTTCCGGTACAGAG
...
633 GTGAGAACGAGATATTCATGGGGGAGTGGAGTTCCGCCCTTCCACAGCCGATTC
|-----|
683 688
693 TTAGATCCTGAAGATGTGTGGGTATCGGGAGCCGCTGTGTAACTT

```

```

...
1220 AGTTTGGGGATTACAGAGAGAGTAGTACCCCTTGCTGGCCAGACGTTTGGCAGCTTCC
|-----|
1270 1275
1280 AGATGACCCCCCAGGAGACACCTGTTGAGCCTGGGGAGAGTAGTACCA
...
1598 TGGCTGAGCCCTGGCCAGTACACAGTGCCTACAGGCCAGTGCTTACACTGGCAGACAA
|-----|
1648 1653
1658 GTTGCTTCCCCAGTGAATCTGGGAGAGGGTTCGTGATCTGCCAAC
...
1882 TTCTAGTCTGATCTGCTTACGCGTACGGGCTGTACCCCTATCTTGGCTAGACGCTTACG
|-----|
1932 1937
1942 AGTGACCCGACGAAGCTTGGCCGCCATGGCCCAACTTGTATTGTC
...
-----
1 match found in sequence:
aas03272 : Mouse mIL12Rb1 Tagman probe.
(from "mycobacterieng.seq")
TOIG of: aas03272 check: 5266 from: 1 to: 20
ID AAS03272 standard; DNA; 20 BP.
XX
XX AAS03272:
AC
XX 07-SEP-2001 (first entry)
DT
XX
XX Mouse mIL12Rb1 Tagman probe.
DE
XX
XX Mouse: type-I cytokine receptor; TCCR; T-cell differentiation;
KW Th1; Th2; agonist; antagonist; autoimmune inflammatory disease;
KW allograft rejection; multiple sclerosis; inflammatory bowel disease;
KW insulin-dependent diabetes mellitus; infectious disease;
KW human immunodeficiency virus; allergic disorder; asthma;
KW allergic rhinitis; HIV; probe; mIL12Rb1; ss.
XX
XX Mus musculus.
OS
XX
XX WO200129070-A2.
PN
XX
XX 26-APR-2001.
PD
XX
XX 18-OCT-2000; 2000WO-US28827.
PF
XX
XX 20-OCT-1999; 99US-0160542.
PR
XX
XX (GETH ) GENENTECH INC.
PA
XX De Sauvage FJ, Grewal I, Gurney AL;
PI
XX WPI: 2001-308474/32.
DR
XX
XX Modulating T-cell differentiation and cytokine release profiles into
PT Th1 and Th2 subtypes, for treating immune-related diseases in mammals,
PT by administering modulator of type I cytokine receptor (TCCR)
XX
XX Example 12; Fig 19; 126pp; English.
PS
XX The sequence is a probe used in a Tagman real-time PCR experiment
CC used to demonstrate that mice deficient of type I cytokine receptor,
CC mTCCR, are impaired in their ability to mount a Th1 response. The
CC invention relates to methods of modulating the differentiation of

```

T-cells into the Th2 subtype instead of the Th1 subtype, by administering a modulator of TCCR (e.g. an antagonist) to enhance, stimulate or potentiate T-cell differentiation, or using TCCR polypeptide or its agonists to prevent, inhibit or attenuate T-cell differentiation. Th1 mediated disease in mammal can be treated by administering a TCCR antagonist and Th2 diseases by administering a TCCR agonist. Th1 mediated diseases include allograft rejection and autoimmune inflammatory diseases, such as allergic encephalomyelitis, multiple sclerosis, insulin-dependent diabetes mellitus, autoimmune uveoretinitis, inflammatory bowel disease or autoimmune thyroid disease. Th2-mediated diseases include infectious diseases, such as Leishmania major, Mycobacterium leprae, Candida albicans, Toxoplasma gondii, respiratory syncytial virus and human immunodeficiency virus (HIV) and allergic disorders, such as asthma, allergic rhinitis, dermatitis and vernal conjunctivitis.

Sequence 20 BP; 1 A; 9 C; 6 G; 4 T; 0 other:

AAS03272 Length: 20 March 5, 2002 14:17 Type: N Check: 5266 ..
Found using 'seq2-3' (pappu403.key)

1 CGCCAGCGTCCTCGTGG
5 10

1 match found in sequence:
aas03273 ; Mouse M1L12Rb2 Tagman sense PCR primer.
(from "mycobactereng.seq")
TOIG of: aas03273 check: 6710 from: 1 to: 21

ID AAS03273 standard; DNA; 21 BP.

XX AAS03273;

XX 07-SEP-2001 (first entry)

DE Mouse M1L12Rb2 Tagman sense PCR primer.

XX Mouse; type-I cytokine receptor; TCCR; T-cell differentiation;

KW Th1; Th2; agonist; antagonist; autoimmune inflammatory disease;

KW allograft rejection; multiple sclerosis; inflammatory bowel disease;

KW insulin-dependent diabetes mellitus; infectious disease;

KW human immunodeficiency virus; allergic disorder; asthma;

KW allergic rhinitis; HIV; PCR primer; M1L12Rb2; ss.

XX Mus musculus.

XX WO200129070-A2.

XX 26-APR-2001.

XX 18-OCT-2000; 2000WO-US28827.

XX 20-OCT-1999; 99US-0160542.

XX (GETH) GENENTECH INC.

XX De Sauvage FJ, Grewal I, Gurney AL.

XX WPI; 2001-308474/32.

XX Modulating T-cell differentiation and cytokine release profiles into
PT Th1 and Th2 subtypes, for treating immune-related diseases in mammals,
PT by administering modulator of type I cytokine receptor (TCCR)
XX
XX Example 12; Fig 19; 126pp; English.

XX The sequence is a PCR primer used in a Tagman real-time PCR experiment
CC used to demonstrate that mice deficient of type I cytokine receptor,
CC mTCCR, are impaired in their ability to mount a Th1 response. The
CC invention relates to methods of modulating the differentiation of
CC T-cells into the Th2 subtype instead of the Th1 subtype, by

administering a modulator of TCCR (e.g. an antagonist) to enhance, stimulate or potentiate T-cell differentiation, or using TCCR polypeptide or its agonists to prevent, inhibit or attenuate T-cell differentiation. Th1 mediated disease in mammal can be treated by administering a TCCR antagonist and Th2 diseases by administering a TCCR agonist. Th1 mediated diseases include allograft rejection and autoimmune inflammatory diseases, such as allergic encephalomyelitis, multiple sclerosis, insulin-dependent diabetes mellitus, autoimmune uveoretinitis, inflammatory bowel disease or autoimmune thyroid disease. Th2-mediated diseases include infectious diseases, such as Leishmania major, Mycobacterium leprae, Candida albicans, Toxoplasma gondii, respiratory syncytial virus and human immunodeficiency virus (HIV) and allergic disorders, such as asthma, allergic rhinitis, dermatitis and vernal conjunctivitis.

Sequence 21 BP; 6 A; 6 C; 3 G; 6 T; 0 other:

AAS03273 Length: 21 March 5, 2002 14:17 Type: N Check: 6710 ..
Found using 'seq2-3' (pappu403.key)

1 CAGCATTTGCATCGCTATCA
12 17

4 matches found in sequence:
aas06592 ; Human immunoregulatory protein B7-H1 cDNA sequence.
(from "mycobactereng.seq")
TOIG of: aas06592 check: 9805 from: 1 to: 3616

ID AAS06592 standard; cDNA; 3616 BP.

XX AAS06592;

XX 26-SEP-2001 (first entry)

DE Human immunoregulatory protein B7-H1 cDNA sequence.

XX Human; immunoregulatory protein; B7-H1; co-stimulating T-cell;

KW B-cell antibody-producing response; IgG2a antibody response; APC;

KW immunodeficiency disease; inflammatory disease; autoimmune disease;

KW antigen presenting cell; pathologic cell mediated disease; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 73..945

FT sig_peptide /tag= a /product= "B7-H1 protein"

FT 73..138

FT mat_peptide /tag= b

FT 139..942

FT /*tag= c

XX WO200139722-A2.

XX 07-JUN-2001.

XX 30-NOV-2000; 2000WO-US32583.

XX 30-NOV-1999; 99US-0451291.

XX 28-AUG-2000; 2000US-0649108.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Chen L;

XX WPI; 2001-397926/42.

XX P-PSDB; AAU03559.

XX Novel DNA encoding immunoregulatory molecule B7-H1, is useful for
PT co-stimulating a T cell for augmenting immunoregulation and for
PT controlling pathologic cell mediated conditions

```

XX  Example 2; Fig 1: 85pp; English.
PS
CC  The present sequence encoding for novel human immunoregulatory protein
CC  B7-H1 (hB7-H1) is capable of co-stimulating T-cells. The sequence for
CC  mouse B7-H1 (mB7-H1) is also given (AAU03560). B7-H1 is useful for
CC  co-stimulating T-cells such as helper T-cells that provide helper
CC  activity for B-cell antibody-producing response e.g. IgG2a antibody
CC  response, in a mammal having an immunodeficiency disease, inflammatory
CC  condition or an autoimmune disease, by culturing B7-H1 with the
CC  mammalian T-cells in vitro, or administering B7-H1 or a nucleic acid
CC  encoding B7-H1 to the T-cells, such that the level of CD40 ligand on the
CC  T-cell surface is increased. The method further involves providing a
CC  recombinant cell e.g. an antigen presenting cell (APC) which is the
CC  progeny of a cell obtained from the mammal and has been transfected or
CC  transformed ex vivo with a nucleic acid encoding B7-H1, so that the cell
CC  expresses B7-H1, and administering the cell to the mammal. Prior to
CC  administration, the APC is pulsed with an antigen or an antigenic
CC  peptide. B7-H1 can be used to control pathologic cell mediated
CC  conditions (e.g. those induced by infectious agents such as Mycobacterium
CC  tuberculosis) or other pathologic cell mediated responses such as those
CC  involved in autoimmune diseases (e.g. rheumatoid arthritis).
XX
SQ  Sequence 3616 BP; 1059 A; 726 C; 739 G; 1092 T; 0 other;
AAS06592 Length: 3616 March 5, 2002 14:18 Type: N Check: 9805 ..
Found using 'seq2-3' (pappu403.key)

...

1139  GGGAGCCTGGAGGAGACCTTGATCTTCAAAATGCTGAGGGGCTCATCGAGCGCTGTG
      |-----|
      1189 1194

1199  ACAGGGAGAAAGACTTCTTGAAACAGAGAGCCTCCAGCAATCA

...

1446  GTTGTGATGATTTCTTTGAAGATATTTGTAGTAGATGTTACATTTTGTCCGCAAC
      |-----|
      1496 1501

1506  TAACTGCTGCTTAATGATTTGCTACACTCTAGTAAACATGAG

...

1718  AAATATCAGCTTTACATTAATGTGTGACCTACACACATTAATCTCATTTCCGCTGTA
      |-----|
      1768 1773

1778  CCACCCGTGTGTGATTAACCACTATTTTACCCTCATGCTACAGCTG

...

1928  AAGCAATTTCTTTATTCAAAAACATTTATTAAAGTCCCTTGCAATATCATCGCTGTGC
      |-----|
      1978 1983

1988  CAGCATTTGAATCTACAGATGTGAGCAAGACAAAGTACCTGCTCTC

...

-----
6 matches found in sequence:
aas06593 ; Mouse immunoregulatory protein B7-H1 cDNA sequence.
(from "mycobacterin.seq")
TOIG of: aas06593 check: 9724 from: 1 to: 873

ID  AAS06593 standard; cDNA; 873 BP.
XX
AC  AAS06593;
XX
DT  26-SEP-2001 (first entry)

```

```

XX  Mouse immunoregulatory protein B7-H1 cDNA sequence.
DE
XX
KW  Mouse; immunoregulatory protein; B7-H1; co-stimulating T-cell;
KW  B-cell antibody-producing response; IgG2a antibody response; APC;
KW  immunodeficiency disease; inflammatory disease; autoimmune disease;
KW  antigen presenting cell; pathologic cell mediated disease; ss.
XX
OS  Mus musculus.
XX
FH  Key Location/Qualifiers
FT  CDS 1..873
FT  /tag= a
FT  /product= "B7-H1 protein"
FT  sig_peptide 1..66
FT  /tag= b
FT  mat_peptide 67..870
FT  /tag= c
XX
PD  WO200139722-A2.
XX
PD  07-JUN-2001.
XX
PE  30-NOV-2000; 2000WO-US32583.
XX
PR  30-NOV-1999; 99US-0451291.
PR  28-AUG-2000; 2000US-0649108.
XX
PA  (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI  Chen L;
XX
PI  WPI: 2001-397926/42.
XX
DR  P-PSDB; AAU03560.
XX
PT  Novel DNA encoding immunoregulatory molecule B7-H1, is useful for
PT  co-stimulating a T cell for augmenting immunoregulation and for
PT  controlling pathologic cell mediated conditions -
XX
PS  Example 7; Fig 10; 85pp; English.
XX
CC  The present sequence encoding for novel mouse immunoregulatory protein
CC  B7-H1 (mB7-H1) is capable of co-stimulating T-cells. The sequence for
CC  human B7-H1 (hB7-H1) is also given (AAU03559). B7-H1 is useful for
CC  co-stimulating T-cells such as helper T-cells that provide helper
CC  activity for B-cell antibody-producing response e.g. IgG2a antibody
CC  response, in a mammal having an immunodeficiency disease, inflammatory
CC  condition or an autoimmune disease, by culturing B7-H1 with the
CC  mammalian T-cells in vitro, or administering B7-H1 or a nucleic acid
CC  encoding B7-H1 to the T-cells, such that the level of CD40 ligand on the
CC  T-cell surface is increased. The method further involves providing a
CC  recombinant cell e.g. an antigen presenting cell (APC) which is the
CC  progeny of a cell obtained from the mammal and has been transfected or
CC  transformed ex vivo with a nucleic acid encoding B7-H1, so that the cell
CC  expresses B7-H1, and administering the cell to the mammal. Prior to
CC  administration, the APC is pulsed with an antigen or an antigenic
CC  peptide. B7-H1 can be used to control pathologic cell mediated
CC  conditions (e.g. those induced by infectious agents such as Mycobacterium
CC  tuberculosis) or other pathologic cell mediated responses such as those
CC  involved in autoimmune diseases (e.g. rheumatoid arthritis).
XX
SQ  Sequence 873 BP; 243 A; 206 C; 235 G; 189 T; 0 other;
AAS06593 Length: 873 March 5, 2002 14:18 Type: N Check: 9724 ..
Found using 'seq2-3' (pappu403.key)

1  ATGAGATATTTGCTGGCATTAATTTACAGAGCTGCTGCACTGCTCAAGGCGCTTACT
      |-----|
      51 56

61  ATCAGGCTCCAAAGACTTGATGCTGTGAGTAGTGCAGCAACGTCACGATGAGTGC
      |-----|
      103 108

```

```

121 AGATTCCCTGTAGAACGGAGCTGGACCTGCTTGCCTT
...
257 CGCTCCCAAGACCAAGCTTTGGAAGGAATGCTGCCCTTCAGATCACAGCAGCTCAGC
      |-----|
      307 312
317 TGCAGACGACGAGCGCTTACTCTGTCGATATATCAGCTACGCGTGTGCGGACTCAAGCGCA
      |-----|
      328 333
377 TCACGCT
...
545 CTTCGCCGACAGAGGGAGTCTTCTCAATGTGACCAAGCAGTCTGAGGGTCAAGCCACAG
      |-----|
      595 600
605 CGAATGATGTTTCTACTGTACGTTTGGAGATCACAGCCAGCGCA
...
767 TCCCTCTCTTTTGGAAAAACAAGTAGAGATGCTAGATGTGGAGAAATGTGCGCTTGAAG
      |-----|
      817 822
827 ATACAAGCTCAAAAAACCGAAATGATACACAATTGAGAGACGTA
...
-----
2 matches found in sequence:
aas08981 : Cpg-containing oligonucleotide sequence 1826.
      (from "mycobacterieng.seq")
      TOIG of: aas08981 check: 5568 from: 1 to: 20

ID   AAS08981 standard; DNA; 20 BP.
XX
AC   AAS08981;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Cpg-containing oligonucleotide sequence 1826.
XX
KW   Cpg motif; saponin; innate immune response; natural killer cell response;
KW   viral disease; hepatitis; feline leukaemia virus; influenza; adenovirus;
KW   herpes simplex virus; HSV; papilloma virus; human immunodeficiency virus;
KW   HIV; bacterial disease; mycoplasma; legionella; anthrax; diphtheria; ds;
KW   Lyme disease; tuberculosis; protozoal disease; leishmania; trypanosoma;
KW   parasitic disease; chlamydia; rickettsia; fibrosarcoma; adenocarcinoma;
KW   retinoblastoma; melanoma; leukaemia; Ewing's tumour; Wilm's tumour;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   WO200151083-A2.
XX
PD   19-JUL-2001.
XX
PE   12-JAN-2001; 2001WO-US01046.
XX
PR   13-JAN-2000; 2000US-0175840.
XX
PR   01-MAY-2000; 2000US-0200853.
XX
PR   06-AUG-2000; 2000US-0369941.
XX
PA   (AQUI-) AQUILA BIOPHARMACEUTICALS INC.
XX
PI   Kensil CR;
XX
DR   WPI; 2001-451816/48.

```

```

XX Composition for enhancing innate immune response and treating
PT infections and cancer comprises a saponin and an oligonucleotide
PT comprising at least one unmethylated Cpg dinucleotide
XX
PS Claim 10; Page 40; 49pp; English.
XX
CC The sequence represents a Cpg motif. Compositions comprising a saponin
CC and a Cpg motif containing at least one unmethylated Cpg
CC (cytosine-guanine) dinucleotide can be administered to humans and other
CC mammals to stimulate an innate immune response and enhance a natural
CC killer cell response. They are therefore useful for treating and
CC preventing viral diseases such as those caused by hepatitis, feline
CC leukaemia virus, influenza, adenovirus, herpes simplex virus (HSV),
CC papilloma virus and human immunodeficiency virus (HIV); bacterial
CC diseases such as those caused by mycoplasma, legionella, anthrax,
CC diphtheria, Lyme disease and tuberculosis; protozoal diseases such as
CC those caused by leishmania and trypanosoma; parasitic diseases such as
CC those caused by chlamydia and rickettsia; and cancers such as
CC fibrosarcoma, adenocarcinoma, retinoblastoma, melanoma, leukaemia,
CC Ewing's tumour and Wilm's tumour.
XX
SQ Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;
AAS08981 Length: 20 March 5, 2002 14:18 Type: N Check: 5568 ..
Found using 'seq2-3' (pappu403.key)
1 TCCATGACGTTCCCTGACGCTT
  6 11 15 20
-----
3 matches found in sequence:
aas08982 : Cpg-containing oligonucleotide sequence 2006.
      (from "mycobacterieng.seq")
      TOIG of: aas08982 check: 3528 from: 1 to: 24

ID   AAS08982 standard; DNA; 24 BP.
XX
AC   AAS08982;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Cpg-containing oligonucleotide sequence 2006.
XX
KW   Cpg motif; saponin; innate immune response; natural killer cell response;
KW   viral disease; hepatitis; feline leukaemia virus; influenza; adenovirus;
KW   herpes simplex virus; HSV; papilloma virus; human immunodeficiency virus;
KW   HIV; bacterial disease; mycoplasma; legionella; anthrax; diphtheria; ds;
KW   Lyme disease; tuberculosis; protozoal disease; leishmania; trypanosoma;
KW   parasitic disease; chlamydia; rickettsia; fibrosarcoma; adenocarcinoma;
KW   retinoblastoma; melanoma; leukaemia; Ewing's tumour; Wilm's tumour;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   WO200151083-A2.
XX
PD   19-JUL-2001.
XX
PE   12-JAN-2001; 2001WO-US01046.
XX
PR   13-JAN-2000; 2000US-0175840.
XX
PR   01-MAY-2000; 2000US-0200853.
XX
PR   06-AUG-2000; 2000US-0369941.
XX
PA   (AQUI-) AQUILA BIOPHARMACEUTICALS INC.
XX
PI   Kensil CR;
XX
DR   WPI; 2001-451816/48.
PT Composition for enhancing innate immune response and treating

```

```

PT      infections and cancer comprises a saponin and an oligonucleotide
PT      comprising at least one unmethylated CpG dinucleotide
PS      Claim 10; Page 40; 49pp; English.
XX      The sequence represents a CpG motif. Compositions comprising a saponin
XX      and a CpG motif containing at least one unmethylated CpG
CC      (cytosine-guanine) dinucleotide can be administered to humans and other
CC      mammals to stimulate an innate immune response and enhance a natural
CC      killer cell response. They are therefore useful for treating and
CC      preventing viral diseases such as those caused by hepatitis, feline
CC      leukaemia virus, influenza, adenovirus, herpes simplex virus (HSV),
CC      papilloma virus and human immunodeficiency virus (HIV); bacterial
CC      diseases such as those caused by mycoplasma, legionella, anthrax,
CC      diphtheria, Lyme disease and tuberculosis; protozoal diseases such as
CC      those caused by leishmania and trypanosoma; parasitic diseases such as
CC      those caused by chlamydia and rickettsia; and cancers such as
CC      fibrosarcoma, adenocarcinoma, retinoblastoma, melanoma, leukaemia,
CC      Ewing's tumour and Wilms' tumour.
XX      Sequence 24 BP; 0 A; 4 C; 6 G; 14 T; 0 other;
SQ
AA508982 Length: 24 March 5, 2002 14:18 Type: N Check: 3528 ..
Found using 'seq2-3' (pappu403.key)
-----
1      |-----| |-----| |-----|
      TCGTCGTTTGTGTCGTTTGTGCTT
      3      8 11 16 19 24
3 matches found in sequence:
aat44026 ; Human cytokine beta-13 CDNA (ATCC 97113).
(from "mycobacterieng.seq")
TOIG of: aat44026 check: 2770 from: 1 to: 282
ID      AAT44026 standard; CDNA; 282 BP.
XX      AAT44026;
AC
XX
XX      03-SEP-1997 (first entry)
DT
XX
XX      Human cytokine beta-13 CDNA (ATCC 97113).
DE
XX      Chemokine beta 13; Ck-beta-13; C-C; Cys-Cys subfamily; immune cell;
KW      defence; activation; eosinophil; monocyte; macrophage; T lymphocyte;
KW      T cell; basophil; gene therapy; tumour; cancer; neoplasia; infection;
KW      Kaposi's sarcoma; cirrhosis; osteoarthritis; pulmonary fibrosis;
KW      leukaemia; autoimmune disease; psoriasis; inflammation; allergy;
KW      rheumatoid arthritis; siliocosis; ss.
XX
XX      Homo sapiens.
OS
XX
XX      Homo sapiens.
FH
XX      Key Location/Qualifiers
FT      CDS 1..282
FT      /*tag= a
FT      /transl_except= pos: 127..129, aa: Tyr
FT      sig_peptide 1..84
FT      /*tag= b
FT      mat_peptide 85..279
FT      /*tag= c
XX
XX      WO9639521-A1.
PN
XX      12-DEC-1996.
PD
XX      06-JUN-1995; 95WO-US07294.
XX
XX      06-JUN-1995; 95WO-US07294.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
PA      (SMK ) SMTTHKLINB BECHAM CORP.
PI      Li H, Seibel G;

```

```

XX WP1: 1997-043143/04.
DR P-PSDB: AAM07604.
XX
PT Human chemokine beta-13 - useful for treating solid tumours,
PT leukaemia, infections, autoimmune disease, fibrotic disorders,
PT psoriasis, etc.
XX
PS Claim 1; Page 45; 58pp; English.
XX
XX AAT44026 encodes human chemokine beta-13 (Ck-beta-13), a member of the
CC C-C (Cys-Cys) branch of intererine chemokines. Ck-beta-13 is useful for
CC treating patients lacking chemokine beta-13 by gene therapy. Ck-beta-13
CC stimulates the invasion and activation of host defence cells making it
CC useful for treating solid tumours, e.g. Kaposi's sarcoma, and for
CC enhancing resistance to acute and chronic infections, e.g. mycobacterial
CC infections. The chemokine induces chemotactic migration of monocytes,
CC neutrophils, eosinophils, T lymphocytes, basophils and fibroblasts to
CC sites where they are needed. Eosinophils may be attracted to the site
CC of a parasitic infection to kill parasite larvae. Ck-beta-13 also
CC recruits debris-clearing and connective tissue promoting inflammatory
CC cells, and is therefore used to stimulate wound healing, prevent
CC scarring and treat liver cirrhosis, osteoarthritis and pulmonary
CC fibrosis. Ck-beta-13 may also be used for treating leukaemia, T-cell
CC mediated autoimmune diseases, psoriasis, to regulate haematopoiesis and
CC to inhibit angiogenesis. Ck-beta-13 antagonists inhibit activity of the
CC chemokine which is useful for treating certain autoimmune diseases,
CC atherosclerosis, chronic inflammatory and infective diseases, allergic
CC reactions, rheumatoid arthritis, sllcosis and bone marrow failure.
XX
XX Sequence 282 BP; 54 A; 86 C; 78 G; 64 T; 0 other:
XX
AAT44026 Length: 282 March 5, 2002 14:18 Type: N Check: 2770 ..
Found using 'seq2-3' (pappu403.key)
1 ATGGCTCGCCTACAGACTGCACCTCCTGGTTGTCCTCGTCCTCTGCTGAGCCCTCAAA 1-----1
61 GGAAGTGAAGGCGAGGCCCTACAGGCCCAATGAGACAGACGCTGCTGCGCGATTAC 1-----1
82 87 100 105
121 GTCCGTCACCGTCTGCCCTGCGCGTGGTGAAGA
...
1 match found in sequence:
aat44050 : Human cytokine beta-13 cDNA vector construction primer.
(from "mycobacteriery.seq")
TOIG of: aat44050 check: 2215 from: 1 to: 60
ID AAT44050 standard: DNA; 60 BP.
XX
XX AAT44050:
XX
XX 03-SEP-1997 (first entry)
XX
DE Human cytokine beta-13 cDNA vector construction primer.
XX
XX Chemokine beta 13; Ck-beta-13; C-C; Cys-Cys subfamily; Immune cell;
XX defence; activation; eosinophil; monocyte; macrophage; T lymphocyte;
XX T cell; basophil; gene therapy; tumour; cancer; neoplasia; infection;
XX Kaposi's sarcoma; cirrhosis; osteoarthritis; pulmonary fibrosis;
XX leukaemia; autoimmune disease; psoriasis; inflammation; allergy;
XX rheumatoid arthritis; sllcosis; PCR; polymerase chain reaction;
XX primer; ss.
XX
XX Synthetic.
XX
XX W09639521-A1.
XX

```

```

PD 12-DEC-1996.
XX
XX 06-JUN-1995; 95WO-US07294.
XX
XX 06-JUN-1995; 95WO-US07294.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (SMK) SMITHKLINE BEECHAM CORP.
XX
XX Li H, Seibel G;
XX
XX WPI: 1997-043143/04.
XX
XX Human chemokine beta-13 - useful for treating solid tumours,
XX leukemia, infections, autoimmune disease, fibrotic disorders,
XX psoriasis, etc.
XX
XX Example 2; Page 37; 58pp; English.
XX
XX AAT44049 and AAT44050 are PCR primers used in the construction of a
XX vector for the mammalian (COS cell) expression of human chemokine
XX beta-13 (Ck-beta-13), a member of the C-C (Cys-Cys) branch of Interleukin
XX chemokines. Ck-beta-13 is useful for treating patients lacking
XX Ck-beta-13 by gene therapy. Ck-beta-13 stimulates the invasion and
XX activation of host defence cells making it useful for treating solid
XX tumours, e.g. Kaposi's sarcoma, and for enhancing resistance to acute
XX and chronic infections, e.g. mycobacterial infections. The chemokine
XX induces chemotactic migration of monocytes, neutrophils, eosinophils,
XX T lymphocytes, basophils and fibroblasts to sites where they are needed.
XX Eosinophils may be attracted to the site of a parasitic infection to
XX kill parasitic larvae (e.g. schistosomiasis). Ck-beta-13 also recruits
XX debris-clearing and connective tissue promoting inflammatory
XX cells, and is therefore used to stimulate wound healing, prevent
XX scarring and treat liver cirrhosis, osteoarthritis and pulmonary
XX fibrosis. Ck-beta-13 may also be used for treating leukaemia, T-cell
XX mediated autoimmune diseases, psoriasis, to regulate haematopoiesis and
XX to inhibit angiogenesis. Ck-beta-13 antagonists inhibit activity of the
XX chemokine which is useful for treating certain autoimmune diseases,
XX atherosclerosis, chronic inflammatory and infective diseases, allergic
XX reactions, rheumatoid arthritis, silirosis and bone marrow failure.
XX
XX Sequence 60 BP; 13 A; 10 C; 18 G; 19 T; 0 other;
XX
AAT44050 Length: 60 March 5, 2002 14:18 Type: N Check: 2215
Found using 'seq2-3' (pappu403.key)
1 CGCTCTAGATTAGGATGCTGGAGCTCGATGCGTATGCTGCTACCTATTAGAGAT
25 30
61
-----
1 match found in sequence:
aat44053 : Human chemokine beta-11 coding sequence.
(from "mycobactereng.seq")
TolG of: aat44053 Check: 9732 from: 1 to: 297
ID AAT44053 standard; cDNA; 297 BP.
XX
XX AAT44053;
AC
XX
XX 05-SEP-1997 (first entry)
DT
XX
XX Human chemokine beta-11 coding sequence.
DE
XX
XX Chemokine beta-11; chemokine alpha-1; human; gene therapy; solid tumour;
XX Kaposi's sarcoma; acute infection; chronic infection; connective tissue;
XX mycobacterial infection; eosinophil; schistosomiasis; pulmonary fibrosis;
XX silirosis; parasite; trichinosis; ascariasis; inflammatory cell;
XX wound healing; liver cirrhosis; osteoarthritis; leukaemia; angiogenesis;
XX haematopoiesis; T-cell mediated autoimmune disease; rheumatoid arthritis;
XX psoriasis; atherosclerosis; chronic inflammatory disease.
KW

```

```

KW allergic reaction; bone marrow failure; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX sig_peptide 1..51
XX FT /*tag= a
XX mat_peptide 52..294
XX FT /*tag= b
XX
XX WO9639522-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96WO-US09572.
XX
XX 05-JUN-1995; 95US-0464401.
XX
XX 05-JUN-1995; 95US-0460987.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Li H;
XX
XX WPI: 1997-043144/04.
XX
XX P-PSDB: AAM07605.
XX
XX Human chemokines beta-11 and alpha-1 - useful for treating solid
XX tumours, leukemia, infections, autoimmune disease, fibrotic
XX disorders, psoriasis, etc.
XX
XX Claim 5; Page 52; 67pp; English.
XX
XX AAT44053 and AAT44054 represent the coding sequences for the chemokines
XX of the invention. This sequence encodes the novel chemokine beta-11
XX (CKB11). AAT44054 encodes the chemokine alpha-1 (CKA1). CKB11 is a
XX member of the C-C branch of chemokines, while CKA1 is a C-X-C chemokine.
XX The encoded proteins, and their agonists are useful for treating patients
XX lacking the respective chemokine, especially by gene therapy. The
XX chemokines stimulate the invasion and activation of host defence cells
XX making them useful for treating solid tumours, e.g. Kaposi's sarcoma, and
XX for enhancing resistance to acute and chronic infections, e.g.
XX mycobacterial infections. The chemokines also increase the presence of
XX eosinophils which kill larvae of tissue invading parasites, e.g.
XX schistosomiasis, trichinosis and ascariasis. The proteins are also able
XX to recruit debris-clearing and connective tissue promoting inflammatory
XX cells, and are therefore used to stimulate wound healing, prevent
XX scarring and to treat liver cirrhosis, osteoarthritis and pulmonary
XX fibrosis. Additionally, the polypeptides can be used for treating
XX leukaemia, T-cell mediated autoimmune diseases, psoriasis, to regulate
XX haematopoiesis and to inhibit angiogenesis. Antagonists of the proteins
XX inhibit activity of the chemokine, and are useful for treating certain
XX autoimmune diseases, atherosclerosis, chronic inflammatory and infective
XX diseases, allergic reactions, rheumatoid arthritis, silirosis and bone
XX marrow failure.
XX
XX Sequence 297 BP; 64 A; 97 C; 76 G; 60 T; 0 other;
XX
AAT44053 Length: 297 March 5, 2002 14:18 Type: N Check: 9732
Found using 'seq2-3' (pappu403.key)
...
231 CTGGGTAGACGATCATCAGAGACTGACAGAGACTGCTGACCAAGATGAGCGCGAG
281 286
291 CAGTTAA
-----
1 match found in sequence:
aat44054 : Human chemokine alpha-1 coding sequence.
(from "mycobactereng.seq")
TolG of: aat44054 Check: 8548 from: 1 to: 330

```

```

ID      AAT44054 standard; cDNA: 330 BP.
XX
AC      AAT44054:
XX
DT      03-SEP-1997 (first entry)
XX
DE      Human chemokine alpha-1 coding sequence.
XX
KW      Chemokine beta-11; chemokine alpha-1; human; gene therapy; solid tumour;
KW      Kaposi's sarcoma; acute infection; chronic infection; connective tissue;
KW      mycobacterial infection; eosinophil; schistosomiasis; pulmonary fibrosis;
KW      silicosis; parasite; trichinosis; ascariasis; inflammatory cell;
KW      wound healing; liver cirrhosis; osteoarthritis; leukaemia; angiogenesis;
KW      haematopoiesis; T-cell mediated autoimmune disease; rheumatoid arthritis;
KW      psoriasis; atherosclerosis; chronic inflammatory disease;
KW      allergic reaction; bone marrow failure; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FT      sig_peptide      1..66      location/Qualifiers
FT      mat_peptide      /**tag= a
FT      /tag= b          67..327
PN      MO9639522-A1.
XX
PD      12-DEC-1996.
XX
PF      05-JUN-1996; 96WO-US09572.
XX
PR      05-JUN-1995; 95US-0464401.
PR      05-JUN-1995; 95US-0460987.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Li H;
XX
DR      WPI: 1997-043144/04.
DR      P-PSDB; AAW07606.
XX
PT      Human chemokines beta-11 and alpha-1 - useful for treating solid
PT      tumours, leukaemia, infections, autoimmune disease, fibrotic
XX      disorders, psoriasis, etc.
XX
PS      Claim 5; Page 52; 67pp; English.
XX
CC      AAT44053 and AAT44054 represent the coding sequences for the chemokines
CC      of the invention. This sequence encodes the novel chemokine alpha-1
CC      (CKA1). AAT44053 encodes the chemokine beta-11 (CKB11). CKB11, is a
CC      member of the C-C branch of chemokines, while CKA1, is a C-X-C chemokine.
CC      The encoded proteins, and their agonists are useful for treating patients
CC      lacking the respective chemokine, especially by gene therapy. The
CC      chemokines stimulate the invasion and activation of host defence cells
CC      making them useful for treating solid tumours, e.g. Kaposi's sarcoma, and
CC      for enhancing resistance to acute and chronic infections, e.g.
CC      mycobacterial infections. The chemokines also increase the presence of
CC      eosinophils which kill larvae of tissue invading parasites, e.g.
CC      schistosomiasis, trichinosis and ascariasis. The proteins are also able
CC      to recruit debris-clearing and connective tissue promoting inflammatory
CC      cells, and are therefore used to stimulate wound healing, prevent
CC      scarring and to treat liver cirrhosis, osteoarthritis and pulmonary
CC      fibrosis. Additionally, the polypeptides can be used for treating
CC      leukaemia, T-cell mediated autoimmune diseases, psoriasis, to regulate
CC      haematopoiesis and to inhibit angiogenesis. Antagonists of the proteins
CC      inhibit activity of the chemokine, and are useful for treating certain
CC      autoimmune diseases, atherosclerosis, chronic inflammatory and infective
CC      diseases, allergic reactions, rheumatoid arthritis, silicosis and bone
CC      marrow failure.
XX
SQ      Sequence 330 BP; 96 A; 67 C; 76 G; 91 T; 0 other:

```

```

AAT44054 Length: 330 March 5, 2002 14:18 Type: N Check: 8548
Found using 'seq2-3' (pappu403.key)

...
84      CACAGCTTGAGGTGTAGATGTGTCCAGAGAGCTCACTCTTTATCCCTAGACGCTTCAT
      |-----|
      134 139

144      TGATCGAATTCGAATCTTGCCCGGTGGGATGTGTTCAGAGAAA
      |
      1
-----
1 match found in sequence:
aat44060 : 3' primer #2 for human chemokine beta-11 coding sequence.
(from "mycobacterng.seq")
TOIG of: aat44060 Check: 9922 from: 1 to: 57

ID      AAT44060 standard; cDNA: 57 BP.
XX
AC      AAT44060:
XX
DT      03-SEP-1997 (first entry)
XX
DE      3' primer #2 for human chemokine beta-11 coding sequence.
XX
KW      Chemokine beta-11; chemokine alpha-1; human; gene therapy; solid tumour;
KW      Kaposi's sarcoma; acute infection; chronic infection; connective tissue;
KW      mycobacterial infection; eosinophil; schistosomiasis; pulmonary fibrosis;
KW      silicosis; parasite; trichinosis; ascariasis; inflammatory cell; primer;
KW      wound healing; liver cirrhosis; osteoarthritis; leukaemia; angiogenesis;
KW      haematopoiesis; T-cell mediated autoimmune disease; rheumatoid arthritis;
KW      psoriasis; atherosclerosis; chronic inflammatory disease; amplify; PCR;
KW      allergic reaction; bone marrow failure; polymerase chain reaction; ss.
XX
OS      Synthetic.
XX
PN      MO9639522-A1.
XX
PD      12-DEC-1996.
XX
PF      05-JUN-1996; 96WO-US09572.
XX
PR      05-JUN-1995; 95US-0464401.
PR      05-JUN-1995; 95US-0460987.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Li H;
XX
DR      WPI: 1997-043144/04.
XX
PT      Human chemokines beta-11 and alpha-1 - useful for treating solid
PT      tumours, leukaemia, infections, autoimmune disease, fibrotic
XX      disorders, psoriasis, etc.
XX
PS      Example 3; Page 40; 67pp; English.
XX
CC      AAT44055-T44066 represent amplification primers for the coding sequences
CC      for the chemokines of the invention (see AAT44053 and AAT44054).
CC      AAT44053 encodes the novel chemokine beta-11 (CKB11), and AAT44054
CC      encodes the chemokine alpha-1 (CKA1). CKB11, is a member of the C-C
CC      branch of chemokines, while CKA1, is a C-X-C chemokine. The encoded
CC      proteins, and their agonists are useful for treating patients lacking the
CC      respective chemokine, especially by gene therapy. The chemokines
CC      stimulate the invasion and activation of host defence cells making them
CC      useful for treating solid tumours, e.g. Kaposi's sarcoma, and for
CC      enhancing resistance to acute and chronic infections, e.g. mycobacterial
CC      infections. The chemokines also increase the presence of eosinophils
CC      which kill larvae of tissue invading parasites, e.g. schistosomiasis,
CC      trichinosis and ascariasis. The proteins are also able to recruit
CC      debris-clearing and connective tissue promoting inflammatory cells, and

```

```

CC are therefore used to stimulate wound healing, prevent scarring and to
CC treat liver cirrhosis, osteoarthritis and pulmonary fibrosis.
CC Additionally, the polypeptides can be used for treating leukaemia,
CC T-cell mediated autoimmune diseases, psoriasis, to regulate
CC haematopoiesis and to inhibit angiogenesis. Antagonists of the proteins
CC inhibit activity of the chemokine, and are useful for treating certain
CC autoimmune diseases, atherosclerosis, chronic inflammatory and infective
CC diseases, allergic reactions, rheumatoid arthritis, silicosis and bone
CC marrow failure.
XX
SQ Sequence 57 BP; 12 A; 11 C; 18 G; 16 T; 0 other:
AAAT44060 Length: 57 March 5, 2002 14:18 Type: N Check: 9922 ..
Found using 'seq2-3' (pappu403.key)

1 CGCTCTAGATTAGCGTAGTCTGGACGTCGATGCGTAGTACTGCTGCGAC
25 30

-----
1 match found in sequence:
aat44062 : 3' primer #2 for human chemokine alpha-1 coding sequence.
(from "mycobactereng.seq")
TOIG of: aat44062 check: 381 from: 1 to: 54

ID AAT44062 standard; cDNA; 54 BP.
XX
AC AAT44062;
XX
DT 03-SEP-1997 (first entry)
XX
DE 3' primer #2 for human chemokine alpha-1 coding sequence.
XX
KW Chemokine beta-11; chemokine alpha-1; human; gene therapy; solid tumour;
KW Kaposi's sarcoma; acute infection; chronic infection; connective tissue;
KW mycobacterial infection; eosinophil; schistosomiasis; pulmonary fibrosis;
KW silicosis; parasite; trichinosis; ascariasis; inflammatory cell; primer;
KW wound healing; liver cirrhosis; osteoarthritis; leukaemia; angiogenesis;
KW haematopoiesis; T-cell mediated autoimmune disease; rheumatoid arthritis;
KW psoriasis; atherosclerosis; chronic inflammatory disease; amplify; PCR;
KW allergic reaction; bone marrow failure; polymerase chain reaction; ss.
XX
OS Synthetic.
XX
PN WO9639522-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96WO-US09572.
XX
PR 05-JUN-1995; 95US-0464401.
XX
PR 05-JUN-1995; 95US-0460987.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Li H;
XX
DR WPI: 1997-043144/04.
XX
PT Human chemokines beta-11 and alpha-1 - useful for treating solid
PT tumours, leukaemia, infections, autoimmune disease, fibrotic
PT disorders, psoriasis, etc.
XX
XX
PS Example 4; Page 42; 67pp; English.
XX
CC AAT44055-144066 represent amplification primers for the coding sequences
CC for the chemokines of the invention (see AAT44053 and AAT44054).
CC AAT44053 encodes the novel chemokine beta-11 (CKB11), and AAT44054
CC encodes the chemokine alpha-1 (CKA1). CKB11 is a member of the C-C
CC branch of chemokines, while CKA1 is a C-X-C chemokine. The encoded
CC proteins, and their agonists are useful for treating patients lacking the
CC respective chemokine, especially by gene therapy. The chemokines
CC stimulate the invasion and activation of host defence cells making them

```

```

CC useful for treating solid tumours, e.g. Kaposi's sarcoma, and for
CC enhancing resistance to acute and chronic infections, e.g. mycobacterial
CC infections. The chemokines also increase the presence of eosinophils
CC which kill larvae of tissue invading parasites, e.g. schistosomiasis,
CC trichinosis and ascariasis. The proteins are also able to recruit
CC debris-clearing and connective tissue promoting inflammatory cells, and
CC are therefore used to stimulate wound healing, prevent scarring and to
CC treat liver cirrhosis, osteoarthritis and pulmonary fibrosis.
CC Additionally, the polypeptides can be used for treating leukaemia,
CC T-cell mediated autoimmune diseases, psoriasis, to regulate
CC haematopoiesis and to inhibit angiogenesis. Antagonists of the proteins
CC inhibit activity of the chemokine, and are useful for treating certain
CC autoimmune diseases, atherosclerosis, chronic inflammatory and infective
CC diseases, allergic reactions, rheumatoid arthritis, silicosis and bone
CC marrow failure.
XX
SQ Sequence 54 BP; 10 A; 10 C; 16 G; 18 T; 0 other:
AAAT44062 Length: 54 March 5, 2002 14:18 Type: N Check: 381 ..
Found using 'seq2-3' (pappu403.key)

1 CGCTCTAGATTAGCGTAGTCTGGACGTCGATGCGTAGGAGATCTTCTCTT
25 30

-----
2 matches found in sequence:
aat49100 : Partial DNA clone Ac11#1-62 encoding immunostimulatory peptide.
(from "mycobactereng.seq")
TOIG of: aat49100 check: 2458 from: 1 to: 265

ID AAT49100 standard; DNA; 265 BP.
XX
AC AAT49100;
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone Ac11#1-62 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9700067-A1.
XX
PD 03-JAN-1997.
XX
PE 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI: 1997-077347/07.
XX
PT New immuno-stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX
PS Claim 1; Page 32; 79pp; English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN) gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to

```


CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
CC encoded by this sequence has amino acid similarity to M. tuberculosis
CC MPTC 190.11C cytochrome C oxidase subunit II and M. leprae sequence in
CC the B1551 region.
XX
SQ Sequence 265 BP; 56 A; 75 C; 73 G; 58 T; 3 other;
AAT49100 Length: 265 March 5, 2002 14:18 Type: N Check: 2458 ..
Found using 'seq2-3' (pappu403.key)
...
50 CGCAGTGGCCACGATGANCNCGGTCTGTGATTCNTCGATCTGGAGAGCTTGACC |-----|
100 105
110 GARTTGTTCACCGGGTTAGCATCAGTCACGCTTGAAACAAGA
...
207 CGATACCGCTTCCGCGACGACACGACACCGGAAATTTCCGTCTGTCGACCTTCG |-----|
257 262

8 matches found in sequence:
aat49101: Partial DNA clone Acl1#1-152 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49101 check: 996 from: 1 to: 484
ID AAT49101 standard; DNA: 484 BP.
XX
AC AAT49101;
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone Acl1#1-152 encoding immunostimulatory peptide.
XX
KM Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9700067-A1.
XX
PD 03-JAN-1997.
XX
PF 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI; 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX isolation
XX
PS Claim 1; Page 32-33; 79pp; English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of

CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
CC encoded by this sequence has amino acid similarity to M. avian
CC acetolactate synthase.
XX
SQ Sequence 484 BP; 87 A; 145 C; 150 G; 90 T; 12 other;
AAT49101 Length: 484 March 5, 2002 14:18 Type: N Check: 996 ..
Found using 'seq2-3' (pappu403.key)
...
1 CTGCTACGACCCCGCAAGAAGACTACGACGAGTGGACACAGAAATTCATATGCGCGCTCAT |-----|
8 13 |-----| 52 57
61 CGGAACCGACGCGCCGACGACGCGTTTGGCTGACGACGATGGTGTTCANTTCGCTTAN |-----|
91 96 94 99
121 CGGTGTCNCTGACTGCCNTTACGACCCTGNTGGCCARSTTGGMTGACACACGATTA |-----|
140 145
181 CGTCGATTCGCTGCT
...
197 GCCGACTATGAGTTCAACGCGCGCCATTACGCTGTGCCGTATGCTCGCTGACGCCGCTG |-----|
247 252
257 TTCTACTACACAAAGGCGGCGCTGGCAACAGCGCGCTTACCCGAC
...
341 CCGAGTTACAGCGCGTGTGTCGNCGCGGCTGATCGGCGACAGGCTGGTACGCCGACC |-----|
391 396
401 TCATCTGCTGACGCTTCCAGGACCGAATGCGCATNCGCGGTGCTACTCGACAAGT |-----|
411 416
461 GGACAT
...

11 matches found in sequence:
aat49102: Partial DNA clone Acl1#1-239 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49102 check: 2401 from: 1 to: 513
ID AAT49102 standard; DNA: 513 BP.
XX
AC AAT49102;
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone Acl1#1-239 encoding immunostimulatory peptide.

```

XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
PN
XX  03-JAN-1997.
PD
XX  14-JUN-1996; 96MO-US10375.
PF
XX  15-JUN-1995; 95US-0000254.
PR
XX  (UYVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI: 1997-077347/07.
DR
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1; Page 33; 79pp; English.
XX
XX  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX  Sequence 513 BP; 95 A; 152 C; 164 G; 96 T; 6 other;
SQ
AAT49102 Length: 513 March 5, 2002 14:18 Type: N Check: 2401 ..
Found using 'seq2-3' (pappu403.key)
1 1-----|
   9 14
   GCGGCGCAGACGTCGGAACCTCGCGCCCAATTGGTGTGGGAMCCGGCATCTTCGACGC
61 AACGCGTTGCGGGTCTTGGCAGTGTTCGATGCCAATCTGCCGCGCGGCGACTGCGCGAT
   |-----|
   108 113
121 GCGGCGCGTTCAAGGAGGCTGTGACAGACCTGGCGTTGCTTCGCGGACACTACTGCC
   |-----|
   154 159
   160 165
181 NAGTGTGTCAGGACCGCTCAAGTGTTCAGTA
   |-----|
   292 297
242 ATCCCAATGTACGGGGTGACAAANNNAATTCGCCGATGTGTGACACGACGTTGACC
   |-----|
302 AATCCCAAGGGCTGACCCACCAATCCGCAATTGCGCGTTCTGTGGATGACAGCGGAAA

```

```

362 ACCGACTTCGCGGATTTTCGTGTGTGGTGGCGGAGACAGTGGCGGGGAGTGTGGCTACGA
   |-----|
   386 391
   392 397
422 ATTCGGGCTCGAGACGTCCTGCTTACAAACCGCGTTCGGCGGCGATGATGCCAATCCGG
   |-----|
   434 439
   452 457
   469 474
482 GGTGTTCAATCAACGAGCGCGCTGGGTACGCG
   |-----|
11 matches found in sequence:
aat49103 : Partial DNA clone AcII#1-247 encoding immunostimulatory peptide.
(from "mycobacterng.seq")
TORG of: aat49103 check: 1663 from: 1 to: 510
-----
ID  AAT49103 standard; DNA; 510 BP.
XX
XX  AAT49103;
AC
XX  23-SEP-1997 (first entry)
DT
XX
XX  Partial DNA clone AcII#1-247 encoding immunostimulatory peptide.
DE
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
PN
XX  03-JAN-1997.
PD
XX  14-JUN-1996; 96MO-US10375.
PF
XX  15-JUN-1995; 95US-0000254.
PR
XX  (UYVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI: 1997-077347/07.
DR
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1; Page 34; 79pp; English.
XX
XX  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells. The protein
CC  encoded by this sequence has amino acid similarity to a peptide
CC  synthetase.
XX

```

SQ Sequence 510 BP; 84 A; 151 C; 164 G; 100 T; 11 other;
 AAT49103 Length: 510 March 5, 2002 14:18 Type: N Check: 1663
 Found using 'seq2-3' (pappu403.key)

```

...
105 TTGCACTGTTGNGAGCCGAGATTGGAAGTTTCCGACATTACCGTTGCGAGCTTGCC
      |-----|
      155 160

165 TCGCCGAGCTTCGCCAAGCCAGGTTGGGACACGCGGTGATTGTGCGTGGGCAATGA
      |-----|
      170 175

225 CGGGCTGCTGGCCCGCCGGAATTCACAGGCTTCGACAGGTTTCCAGACCGGGC
      |-----|
      252 257

285 CGCTCGCTAGCGGTGACGAGTCTTCTGTAATTCGGCGATCAGCAGTGAACCTACCGCA
      |-----|
      319 324

345 CCGTAAGCGCCACCGCCACCGGTNNCGCGGTGTCGCGCANNCGCGGCGTGGCCCGG
      |-----|
      349 354
      391 396
      403

405 CGACGTCGTTGGCATCATGTTGCGTAATCACCAGCAGCAGCTTGGCGTGGCCAC
      |-----|
      406 411
      409 414
      450 455

465 GGTCAGTGGCGCGTATCGCGCATGCTCAACTACACACGACGCGG
      |-----|
      480 485

7 matches found in sequence:
aat49104 ; Partial DNA clone Acii#1-426 encoding immunostimulatory peptide.
(from "Mycobacterieng.seq")
TOIG of: aat49104 check: 846 from: 1 to: 456

ID AAT49104 standard; DNA: 456 BP.
XX
AC AAT49104;
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone Acii#1-426 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
FN W09700067-A1.
XX
PD 03-JAN-1997.
XX
PE 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI: 1997-077347/07.
XX

```

```

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
PS Claim 1; Page 34; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX which encode partial sequences of immunostimulatory peptides. Each of
XX the clones encode at least one immunostimulatory T cell epitope. The
XX clones were identified by testing over 300 fusion clones (alkaline
XX phosphatase-M. tuberculosis peptide fusions) for their ability to
XX stimulate interferon (IFN)-gamma production. 80 clones were initially
XX designated to have some ability to stimulate IFN-gamma production, of
XX which 76 are shown in AAT49100-175. These sequences can be used to
XX obtain the full length M. tuberculosis genes and corresponding proteins
XX using standard techniques. The peptides are useful in vaccines, as
XX reagents in an improved tuberculin skin test (especially using peptides
XX different from those used in vaccines so as to allow differentiation
XX between vaccinated and infected subjects) and as immunoassay reagents
XX for detecting specific antibodies. An advantage of these peptides is
XX that they stimulate production of IFN-gamma (critical for a protective
XX immune response to M. tuberculosis) by CD4-positive T cells. The protein
XX encoded by this sequence has amino acid similarity to a dipeptide
XX transport protein.
SQ Sequence 456 BP; 63 A; 121 C; 165 G; 98 T; 9 other;

AAT49104 Length: 456 March 5, 2002 14:18 Type: N Check: 846
Found using 'seq2-3' (pappu403.key)

...
130 GCGCTGATGGCGGAGCAKTTGTGACCGAAGANTCTTCAATCCAGCGCTGGGG
      |-----|
      180 185

190 GCGTGTGTATCAGCGGTGACCGTACAGAGACGCGGNGTGTGATGATGACGCT
      |-----|
      220 225

250 GCTGTGCTGATCTACCTGATCACCACATCTGTGTGATGCTGTATTCGCGCCCTGGA
      |-----|
      308 308

310 GCGCGNMGATNCGCTATGCTGAGCAGCGGGTTCTGCTGATGCTTCCGCGGCTTG
      |-----|
      313

370 GCGCGGCTGCTCAANTGCTGATGCGCGGCGTGAACCTGATTTCTTGTGCGGGG
      |-----|
      374 379
      398 403
      428

430 GGTTCGCTGCTGTTGTTACCGCAGCGG
      |-----|
      433 437
      442

4 matches found in sequence:
aat49106 ; Partial DNA clone Acii#2-23 encoding immunostimulatory peptide.
(from "Mycobacterieng.seq")
TOIG of: aat49106 check: 7655 from: 1 to: 231

ID AAT49106 standard; DNA: 231 BP.
XX
AC AAT49106;
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone Acii#2-23 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX tuberculin skin test; ds.
XX

```

```

XX OS Mycobacterium tuberculosis.
XX PN W09700067-A1.
XX PD 03-JAN-1997.
XX PE 14-JUN-1996; 96WO-US10375.
XX PR 15-JUN-1995; 95US-0000254.
XX PA (UYVI-) UNIV VICTORIA.
XX PI Nano FE;
XX DR WPI; 1997-077347/07.
XX DR
XX PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
XX PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX PT isolation
XX PS Claim 1; Page 35; 79pp; English.
XX
CC AA49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AA49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence has homology to a region within sequence MD0009 of the M.
CC leprae chromosome.
XX
SQ Sequence 231 BP; 34 A; 79 C; 79 G; 36 T; 3 other;
AA49106 Length: 231 March 5, 2002 14:18 Type: N Check: 7655
Found using 'seq2-3' (pappu403.key)
1 CTTCTCGCGCCAGCCGTCCTCCGCTGTCGGGATGCGCTACCGGTCAGCGCCCAAGACGG
|-----|-----|
42 47 48 53
61 TGCAGCTCAGACGAGCGGGTGTGGTGGCAGGTCGACTTCCGCGCCCAATGTSGCGG
|-----|
69 74
121 GGCT
....
140 GTGCCGCTGTGTCAAANNNGCAGACGTCGTGTCGCCGCGACGCGCCCGATCGTCGAG
|-----|
190 195
200 GCATGCAGATCCAGTGATCCCGCAATCGAT
-----
5 matches found in sequence:
aat49107; Partial DNA clone AcII#2-26 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49107 check: 3318 from: 1 to: 173

```

```

ID AA49107 standard; DNA; 173 BP.
XX
AC AA49107;
XX
XX 23-SEP-1997 (first entry)
XX
DE Partial DNA clone AcII#2-26 encoding immunostimulatory peptide.
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
XX KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX KW tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX W09700067-A1.
XX PN
XX PD 03-JAN-1997.
XX PF 14-JUN-1996; 96WO-US10375.
XX PR 15-JUN-1995; 95US-0000254.
XX PA (UYVI-) UNIV VICTORIA.
XX PI Nano FE;
XX PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
XX PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX PT isolation
XX PS Claim 1; Page 36; 79pp; English.
XX
CC AA49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AA49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
CC
SQ Sequence 173 BP; 38 A; 51 C; 44 G; 37 T; 3 other;
AA49107 Length: 173 March 5, 2002 14:18 Type: N Check: 3318
Found using 'seq2-3' (pappu403.key)
1 GTTCGCGCGCTCAAAAGATTGACGATGTCACGTCGACGTCGTGCGCCGAGACCAAGT
|-----|
22 27
61 GGAATTGCGTGAAGACCTCAAGAANCCTACTCGNATGTCAGAGCCCTCAACGACGAC
|-----|
96 101 111 116 114 119
121 GAAAGATTTCGTCACCTCGTCGACGTCGTGTCGACGTCCTCCCATTTCCCAAC
|-----|
155 160
-----
2 matches found in sequence:
aat49108; Partial DNA clone AcII#2-35 encoding immunostimulatory peptide.

```


2 matches found in sequence:

aat49110 ; Partial DNA clone AcII#2-506 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49110 check: 5269 from: 1 to: 160

```
ID  AAT49110 standard; DNA; 160 BP.
XX
XX  AAT49110;
AC
XX  23-SEP-1997 (first entry)
DT
XX
XX  Partial DNA clone AcII#2-506 encoding immunostimulatory peptide.
DE
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
PN
XX  03-JAN-1997.
PD
XX  14-JUN-1996; 96WO-US10375.
PF
XX  15-JUN-1995; 95US-0000254.
PR
XX  (UYVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI; 1997-077347/07.
DR
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1; Page 37; 79pp; English.
XX
CC  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ  Sequence 160 BP; 24 A; 52 C; 45 G; 17 T; 22 other;

```

AAT49110 Length: 160 March 5, 2002 14:18 Type: N Check: 5269 ..
Found using 'seq2-3' (pappu403.key)

```
1  CNGGCNNCCAMCGGTCGCCACGACGCGCGTCCTGCGAGATTCGTCNCNAATGGCAN
      |-----|
      44      49

```

```
61  GCGGCGCCCAKANNANNCCGATACGCTTGCTGTTGNGWCAMCCTTNGGACCAACCC
      |-----|
      64      69

```

3 matches found in sequence:

aat49111 ; Partial DNA clone AcII#2-508 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49111 check: 2585 from: 1 to: 133

```
ID  AAT49111 standard; DNA; 133 BP.
XX
XX  AAT49111;
AC
XX  23-SEP-1997 (first entry)
DT
XX
XX  Partial DNA clone AcII#2-508 encoding immunostimulatory peptide.
DE
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
PN
XX  03-JAN-1997.
PD
XX  14-JUN-1996; 96WO-US10375.
PF
XX  15-JUN-1995; 95US-0000254.
PR
XX  (UYVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI; 1997-077347/07.
DR
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1; Page 37; 79pp; English.
XX
CC  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells. The protein
CC  encoded by this sequence has amino acid similarity to a sequence of the
CC  M. leprae ORF encoding gp AAU00018 and similar to a sequence in the B2168
CC  c2-209 region of the M. leprae genome.
XX
SQ  Sequence 133 BP; 14 A; 34 C; 45 G; 28 T; 12 other;

```

AAT49111 Length: 133 March 5, 2002 14:18 Type: N Check: 2585 ..
Found using 'seq2-3' (pappu403.key)

```
29  THGNGHGCHGCANTGCATCTGCGCTTGCTTGCGGCGTTGTCATTCGCTTCG
      |-----|
      79      84

```

```
89  ACCAGTTGTGCGCGCTGAGACAGCATATGCGGCGCTGTGCTATTCG
      |-----|
      98     103     117    122

```

6 matches found in sequence:

aat49112 : Partial DNA clone AcI1#2-511 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49112 check: 7714 from: 1 to: 421

```
ID  AAT49112 standard; DNA: 421 BP.
XX
XX  AAT49112:
AC
XX
XX  23-SEP-1997 (first entry)
DT
XX
XX  Partial DNA clone AcI1#2-511 encoding immunostimulatory peptide.
DE
XX
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
PN
XX  03-JAN-1997.
PD
XX  14-JUN-1996; 96WO-US10375.
PF
XX  15-JUN-1995; 95US-0000254.
PR
XX  (UYVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI; 1997-077347/07.
DR
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1; Page 38; 79pp: English.
PS
XX  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX  Sequence 421 BP: 64 A; 107 C; 145 G; 66 T; 39 other:
SQ

AAT49112 Length: 421 March 5, 2002 14:18 Type: N Check: 7714 ..
Found using 'seq2-3' (pappu403.key)

1      GCGNACNCTGCGCATNCTGCGCNCSTANCCGCGCCGCGCATGAGCNCNTAGCGCGAAG
      |-----|
      31      36

61     TGCGTGTGTTKANCNTNCGCTGTGTGTG

...

140    AGGTGACGGGTGGCGCGTNAAGTGCAGAGTTNGASNTCTGCGCGTGTGCGCCTTNGGT
      |-----|
      190    195
```

2 matches found in sequence:

aat49113 : Partial DNA clone AcI1#2-523 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49113 check: 9110 from: 1 to: 175

```
ID  AAT49113 standard; DNA: 175 BP.
XX
XX  AAT49113:
AC
XX
XX  23-SEP-1997 (first entry)
DT
XX
XX  Partial DNA clone AcI1#2-523 encoding immunostimulatory peptide.
DE
XX
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
PN
XX  03-JAN-1997.
PD
XX  14-JUN-1996; 96WO-US10375.
PF
XX  15-JUN-1995; 95US-0000254.
PR
XX  (UYVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI; 1997-077347/07.
DR
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1; Page 38; 79pp: English.
PS
XX  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
```

between vaccinated and infected subjects) and as immunoassay reagents

the clones encode at least one immunostimulatory T cell epitope. T

CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in T49100-175. These sequences can be used to obtain
CC the full length M. tuberculosis genes and corresponding proteins using
CC standard techniques. The peptides are useful in vaccines, as reagents in
CC an improved tuberculin skin test (especially using peptides different
CC from those used in vaccines so as to allow differentiation between
CC vaccinated and infected subjects) and as immunoassay reagents for
CC detecting specific antibodies. An advantage of these peptides is that
CC they stimulate production of IFN-gamma (critical for a protective immune
CC response to M. tuberculosis) by CD4-positive T cells. This sequence has
CC homology to a M. tuberculosis sequence within the region MD0074
CC (Genbank reference U27357).
XX
SQ Sequence 168 BP; 32 A; 59 C; 40 G; 22 T; 15 other;
AAT49115 Length: 168 March 5, 2002 14:19 Type: N Check: 2012 ..
Found using 'seq2-3' (pappu403.key)
...
9 GAAGCGCACCCCGGTCCGGAAGNCCTGACAGATTCCWNCGTGCTCGCGGACGCTACATC |-----|
59 64
69 CTCGNTGCGGCGGTCCCAAGCAGCCAGCAGCATCATCTTACCT
...
1 match found in sequence:
aat49116 : Partial DNA clone AcII#2-854 encoding immunostimulatory peptide.
(from "Mycobactering.seq")
TOIG of: aat49116 check: 6419 from: 1 to: 181

ID AAT49116 standard; DNA; 181 BP.
XX
AC AAT49116.
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone AcII#2-854 encoding immunostimulatory peptide.
XX
KM Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9700067-A1.
XX
PD 03-JAN-1997.
XX
PF 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI; 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX
PS Claim 1; Page 40; 79pp: English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The

CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ Sequence 181 BP; 28 A; 60 C; 55 G; 28 T; 10 other;
AAT49116 Length: 181 March 5, 2002 14:19 Type: N Check: 6419 ..
Found using 'seq2-3' (pappu403.key)
...
10 CGCCGNCNTNACNCGGTGCGTGCACCGGCCANCTGACGCTGCCCGACGCGAAG |-----|
60 65
70 TGGTGTTCGNCGCCGCGGCGCTTCGAACCGCGCGGATGGCAACG
...
1 match found in sequence:
aat49117 : Partial DNA clone AcII#2-872 encoding immunostimulatory peptide.
(from "Mycobactering.seq")
TOIG of: aat49117 check: 9039 from: 1 to: 95

ID AAT49117 standard; DNA; 95 BP.
XX
AC AAT49117.
XX
DT 23-SEP-1997 (first entry)
XX
DE Partial DNA clone AcII#2-872 encoding immunostimulatory peptide.
XX
KM Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9700067-A1.
XX
PD 03-JAN-1997.
XX
PF 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI; 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
XX
PS Claim 1; Page 40; 79pp: English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline

CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.
 XX
 SQ Sequence 95 BP; 16 A; 34 C; 29 G; 11 T; 5 other;
 AAT49117 Length: 95 March 5, 2002 14:19 Type: N Check: 9039 ..
 Found using 'seq2-3' (pappu403.key)

1 AGTACGCGTGGAGCGCCGCCGTCGATCGGTCGAGAGATGCTCCGACAGCGCA
 14 19

61 GCTGCGTCC

4 matches found in sequence:
 aat49119 ; Partial DNA clone Acil#2-8841 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49119 check: 8746 from: 1 to: 156

ID AAT49119 standard; DNA: 156 BP.
 XX AAT49119;
 AC
 XX 23-SEP-1997 (first entry)
 DT
 XX
 DE Partial DNA clone Acil#2-8841 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 OS
 XX Mycobacterium tuberculosis.
 XX
 PN WO9700067-A1.
 PD
 XX 03-JAN-1997.
 PF 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI; 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 41; 79pp; English.
 XX
 CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of

CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells. The protein
 CC encoded by this sequence has amino acid similarity to isocitrate
 CC dehydrogenase.
 XX
 SQ Sequence 156 BP; 37 A; 48 C; 38 G; 28 T; 5 other;
 AAT49119 Length: 156 March 5, 2002 14:19 Type: N Check: 8746 ..
 Found using 'seq2-3' (pappu403.key)

1 TCTTTACAGAGAGCGCTTTCGCCAGACACGAGACTGTTCCAGCATTTGGNCGTCAAG
 12 17 42 47 57

61 TCAACATGCGCTTTCGATCTGTACRAGCAGTTCGAGTCCGCCGBCGACAGCGCA
 62 99 104

121 CGAGATCATCGAGAGACCTACACCGTTGCCACGAA

1 match found in sequence:
 aat49120 ; Partial DNA clone Acil#2-8941 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49120 check: 1942 from: 1 to: 123

ID AAT49120 standard; DNA: 123 BP.
 XX AAT49120;
 AC
 XX 26-SEP-1997 (first entry)
 DT
 XX
 DE Partial DNA clone Acil#2-8941 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 OS
 XX Mycobacterium tuberculosis.
 XX
 PN WO9700067-A1.
 PD
 XX 03-JAN-1997.
 PF 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI; 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 41; 79pp; English.
 XX
 CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline

CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.

SQ Sequence 123 BP; 11 A; 39 C; 40 G; 24 T; 9 other;

AAT49120 Length: 123 March 5, 2002 14:19 Type: N Check: 1942 ..
 Found using 'seq2-3' (pappu403.key)

42 CGGTGACCCAGTGGCGGCTGCTGACGCGAGTGTGNCAGCCTCTCG |-----|
 92 97

102 GTGCGTGNCTNTCTCCGCC

 1 match found in sequence:
 aat49122 : Partial DNA clone Acl1#2-916 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49122 check: 1956 from: 1 to: 103

ID AAT49122 standard; DNA: 103 BP.

AC AAT49122;

DT 26-SEP-1997 (first entry).

DE Partial DNA clone Acl1#2-916 encoding immunostimulatory peptide.

XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

XX Mycobacterium tuberculosis.

OS WO9700067-A1.

PN 03-JAN-1997.

PD 14-JUN-1996; 96MO-US10375.

PF 15-JUN-1995; 95US-0000254.

PR (UYVI-) UNIV VICTORIA.

PA Nano FE;

PI WPI; 1997-077347/07.

DR New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

XX useful in vaccines, diagnostic skin test, immunoassay and gene

PT isolation

PS Claim 1; Page 42; 79pp; English.

XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of

CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.

SQ Sequence 103 BP; 16 A; 45 C; 27 G; 15 T; 0 other;

AAT49122 Length: 103 March 5, 2002 14:19 Type: N Check: 1956 ..
 Found using 'seq2-3' (pappu403.key)

1 CTTCCGGCGGACACACAGTCTCACCGGCGCACCCCTGACACCTGATCGCTCTG |-----|
 30 35

61 CCGATCCCGGTGCGAGCACCGGGGT

 2 matches found in sequence:
 aat49123 : Partial DNA clone Acl1#2-1014 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49123 check: 459 from: 1 to: 207

ID AAT49123 standard; DNA: 207 BP.

AC AAT49123;

DT 26-SEP-1997 (first entry)

DE Partial DNA clone Acl1#2-1014 encoding immunostimulatory peptide.

XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

XX Mycobacterium tuberculosis.

OS WO9700067-A1.

PN 03-JAN-1997.

PD 14-JUN-1996; 96MO-US10375.

PF 15-JUN-1995; 95US-0000254.

PR (UYVI-) UNIV VICTORIA.

PA Nano FE;

PI WPI; 1997-077347/07.

DR New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

XX useful in vaccines, diagnostic skin test, immunoassay and gene

PT isolation

PS Claim 1; Page 43; 79pp; English.

XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as

CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.
 XX
 SQ Sequence 207 BP; 28 A; 82 C; 59 G; 29 T; 9 other;

AAT49123 Length: 207 March 5, 2002 14:19 Type: N Check: 459 ..
 Found using 'seq2-3' (pappu403.key)

41 GCCTCAGGGGATCCCTGCTGCCACCGCCCTATCTCCGCCAGTCTGGCGGCA |-----|
 91 96
 101 ACAGTCCCGYCCGTGCAGACCTCAGCGCTNCTHCCGGGGAGCAGACGAGTTCTC
 161 ACCGGYCCMNNGYACCCGACCGATCGCGTCCGATTCGGTGGCA
 189 194

 1 match found in sequence:
 aat49124 : Partial DNA clone AcII#2-1025 encoding immunostimulatory peptide.
 (from "Mycobacterieng.seq")
 TOIG of: aat49124 check: 9014 from: 1 to: 204

ID AAT49124 standard; DNA: 204 BP.
 XX
 AC AAT49124;
 XX
 DT 26-SEP-1997 (first entry)
 XX
 DE Partial DNA clone AcII#2-1025 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09700067-A1.
 XX
 PD 03-JAN-1997.
 XX
 PF 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI: 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 43; 79pp; English.
 XX
 CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins

CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.
 XX
 SQ Sequence 204 BP; 29 A; 63 C; 53 G; 45 T; 14 other;

AAT49124 Length: 204 March 5, 2002 14:19 Type: N Check: 9014 ..
 Found using 'seq2-3' (pappu403.key)

93 TGTGTACTNAGAACCTGATCGCGCATTTTCACACACCGTNGCAGACATCGCACCG |-----|
 143 148
 153 TCCGGCCGCYCGCGTGGAGTGTGATCTTGTCGAGTAGAG

 1 match found in sequence:
 aat49125 : Partial DNA clone AcII#2-1035 encoding immunostimulatory peptide.
 (from "Mycobacterieng.seq")
 TOIG of: aat49125 check: 4369 from: 1 to: 207

ID AAT49125 standard; DNA: 207 BP.
 XX
 AC AAT49125;
 XX
 DT 26-SEP-1997 (first entry)
 XX
 DE Partial DNA clone AcII#2-1035 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09700067-A1.
 XX
 PD 03-JAN-1997.
 XX
 PF 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI: 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 43; 79pp; English.
 XX
 CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as

CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is that
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.

Sequence 207 BP; 23 A; 77 C; 57 G; 41 T; 9 other;

AA149125 Length: 207 March 5, 2002 14:19 Type: N Check: 4369 ..
 Found using 'seq2-3' (pappu403.key)

86 NCCTTCACGGCTTGCCGCTTCGGGTAGAGACCGTGCATGCCCGCTTGGGCGCTCGG
 136 141

146 GTGTGCGTCTGCTGCGGCTGCTGCGCNCCTGCGCCACCC

3 matches found in sequence:
 aat49126 ; Partial DNA clone Ac11#2-1084 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49126 check: 6696 from: 1 to: 289

ID AAT49126 standard; DNA: 289 BP.

AC AAT49126;

DT 26-SEP-1997 (first entry)

DE Partial DNA clone Ac11#2-1084 encoding immunostimulatory peptide.

KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KM stimulate; interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

OS Mycobacterium tuberculosis.

PN WO9700067-A1.

PD 03-JAN-1997.

PF 14-JUN-1996; 96WO-US10375.

PR 15-JUN-1995; 95US-0000254.

PA (UYVI-) UNIV VICTORIA.

Nano FE;

WPI; 1997-077347/07.

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

PT useful in vaccines, diagnostic skin test, immunoassay and gene

PT isolation

PS Claim 1; Page 44; 79pp; English.

CC T49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in T49100-175. These sequences can be used to obtain
 CC the full length M. tuberculosis genes and corresponding proteins using
 CC standard techniques. The peptides are useful in vaccines, as reagents in
 CC an improved tuberculin skin test (especially using peptides different

CC from those used in vaccines so as to allow differentiation between
 CC vaccinated and infected subjects) and as immunoassay reagents for
 CC detecting specific antibodies. An advantage of these peptides is that
 CC they stimulate production of IFN-gamma (critical for a protective immune
 CC response to M. tuberculosis) by CD4-positive T cells. This sequence has
 CC homology to sequences within M. tuberculosis clone X68281 (Genbank
 CC Reference Number) and M. leprae clone B983.

Sequence 289 BP; 41 A; 83 C; 86 G; 55 T; 24 other;

AA149126 Length: 289 March 5, 2002 14:19 Type: N Check: 6696 ..
 Found using 'seq2-3' (pappu403.key)

134 ACTTCACCTGCTGTCGGGCTTATCCCGGTGACCAAGTCCGCTGCTGCGCTCCNGG
 184 189

194 TGGGTGAGATCGACATGATGAGCCGCGGTCCGACGTSAGATCATATGCGGTGT
 222 227

254 CCAAGGACGTCAAGTGCCTGCGTGSACGNTGCAGGCC
 259 264

3 matches found in sequence:
 aat49127 ; Partial DNA clone Ac11#2-1089 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49127 check: 9239 from: 1 to: 198

ID AAT49127 standard; DNA: 198 BP.

AC AAT49127;

DT 26-SEP-1997 (first entry)

DE Partial DNA clone Ac11#2-1089 encoding immunostimulatory peptide.

KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KM stimulate; interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

OS Mycobacterium tuberculosis.

PN WO9700067-A1.

PD 03-JAN-1997.

PF 14-JUN-1996; 96WO-US10375.

PR 15-JUN-1995; 95US-0000254.

PA (UYVI-) UNIV VICTORIA.

Nano FE;

WPI; 1997-077347/07.

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

PT useful in vaccines, diagnostic skin test, immunoassay and gene

PT isolation

PS Claim 1; Page 44; 79pp; English.

CC T49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially

CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in T49100-175. These sequences can be used to obtain
 CC the full length M. tuberculosis genes and corresponding proteins using
 CC standard techniques. The peptides are useful in vaccines, as reagents in
 CC an improved tuberculin skin test (especially using peptides different
 CC from those used in vaccines so as to allow differentiation between
 CC vaccinated and infected subjects) and as immunoassay reagents for
 CC detecting specific antibodies. An advantage of these peptides is that
 CC they stimulate production of IFN-gamma (critical for a protective immune
 CC response to M. tuberculosis) by CD4-positive T cells. This sequence has
 CC homology to sequences within M. tuberculosis clone X73561 (Genbank
 CC Reference Number) and M. bovis MD0057 and U34849 regions. The protein
 CC encoded by Act1#2-1089 has homology to immunogenic proteins MPB64 and
 CC MPB64.

SQ Sequence 198 BP; 48 A; 55 C; 52 G; 39 T; 4 other;
 AAT49127 Length: 198 March 5, 2002 14:19 Type: N Check: 9239 ..
 Found using 'seq2-3' (pappu403.key)

20 AGCAATCACCCTATGACACGCTGTGGAGGCTGACACCGATCCGCTGCCATCGCTCC
 70 75
 80 CCATTGTGCAAGTGTGACTGAGCAANGCAGACCGACACAGATGATCGATAGCGCGGAAT
 130 135
 140 GCCGGCTTGACCCGGTGAATATCAGAACTTTCAGTACGACGACGCGGTGATTTT
 182 187

 1 match found in sequence:
 aat49129; Partial DNA clone Act1#2-1104 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49129 check: 5059 from: 1 to: 210

ID AAT49129 standard; DNA; 210 BP.
 AC AAT49129;
 XX
 DT 26-SEP-1997 (first entry)
 DE Partial DNA clone Act1#2-1104 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 OS
 XX Mycobacterium tuberculosis.
 PN WO9700067-A1.
 XX
 PD 03-JAN-1997.
 XX
 PF 14-JUN-1996; 96MO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI; 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 45; 79pp; English.

XX T49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN) gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in T49100-175. These sequences can be used to obtain
 CC the full length M. tuberculosis genes and corresponding proteins using
 CC standard techniques. The peptides are useful in vaccines, as reagents in
 CC an improved tuberculin skin test (especially using peptides different
 CC from those used in vaccines so as to allow differentiation between
 CC vaccinated and infected subjects) and as immunoassay reagents for
 CC detecting specific antibodies. An advantage of these peptides is that
 CC they stimulate production of IFN-gamma (critical for a protective immune
 CC response to M. tuberculosis) by CD4-positive T cells. This sequence has
 CC homology to sequences within M. tuberculosis clone X80268 (Genbank
 CC Reference Number). The protein encoded by it has homology to Cds1 in
 CC M. leprae sequence region MD0045 and to secreted antigenic protein.

SQ Sequence 210 BP; 42 A; 78 C; 59 G; 31 T; 0 other;
 AAT49129 Length: 210 March 5, 2002 14:19 Type: N Check: 5059 ..
 Found using 'seq2-3' (pappu403.key)

53 AGCAGCCGCGGTATCGTCTCGGCCGCTTAGACCAAAACTTACGCCAGCGCCGCA
 103 108
 113 GCCACCCGACTCCAGAGCCCTCGGCCGCTTGGGTTCCGACATGGGT

 1 match found in sequence:
 aat49130; Partial DNA clone Act1#3-9 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49130 check: 7536 from: 1 to: 255

ID AAT49130 standard; DNA; 255 BP.
 AC AAT49130;
 XX
 DT 26-SEP-1997 (first entry)
 DE Partial DNA clone Act1#3-9 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 OS
 XX Mycobacterium tuberculosis.
 PN WO9700067-A1.
 XX
 PD 03-JAN-1997.
 XX
 PF 14-JUN-1996; 96MO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI; 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation

XX Claim 1, Page 46; 79pp; English.
PS
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ Sequence 255 BP; 47 A; 93 C; 69 G; 42 T; 4 other:
AAT49130 Length: 255 March 5, 2002 14:19 Type: N Check: 7536 ..
Found using 'seq2-3' (pappu403.key)
...
38 ACCCTGCACACGTGCGGCTCAACCTGCANTGACGCGAACCTCTGGTGGCGCTCAAC |-----|
88 93
98 GAGTGTTCGGCGGCTCAACCCGCTGACCCCGCGAACCCTGCACGT
...
1 match found in sequence:
aat49131: Partial DNA clone AcII#3-12 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49131 Check: 7570 from: 1 to: 164
ID AAT49131 standard; DNA: 164 BP.
XX
AC AAT49131;
XX
DT 26-SEP-1997 (first entry)
XX
DE Partial DNA clone AcII#3-12 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9700067-A1.
PD 03-JAN-1997.
XX
PF 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNITV VICTORIA.
XX
PI Nano FE;
XX
DR WPI; 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX

PS Claim 1, Page 46; 79pp; English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence has homology to M. leprae sequence within genomic clone B1529,
CC and the protein encoded AcII#3-12 has amino acid similarity to
CC penicillin binding protein.
XX
SQ Sequence 164 BP; 22 A; 53 C; 44 G; 27 T; 18 other:
AAT49131 Length: 164 March 5, 2002 14:19 Type: N Check: 7570 ..
Found using 'seq2-3' (pappu403.key)
...
65 NANTGANTMNCGCGGTGCGGCTCGAGCGTGACCTGGCATCCCTGAGACGCTGCTG |-----|
115 120
125 GGTTCGCCCGGGGAGNTCGAMANTCGGGCATCGCACCATC
...
5 matches found in sequence:
aat49132: Partial DNA clone AcII#3-15 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49132 Check: 7863 from: 1 to: 237
ID AAT49132 standard; DNA: 237 BP.
XX
AC AAT49132;
XX
DT 26-SEP-1997 (first entry)
XX
DE Partial DNA clone AcII#3-15 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9700067-A1.
PD 03-JAN-1997.
XX
PF 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNITV VICTORIA.
XX
PI Nano FE;
XX
DR WPI; 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX


```

PF 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
PR
XX (UYVI-) UNIV VICTORIA.
PA
XX Nano FE;
PI
XX WPI; 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 48; 79pp: English.
XX
XX AAT49136 standard; DNA: 312 BP.
XX
XX AAT49136;
AC
XX 10-NOV-1997 (first entry)
DT
XX Partial DNA clone AclI#3-134 encoding immunostimulatory peptide.
DE
XX Mycobacterium tuberculosis; immunostimulatory peptide: T cell epitope:
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS

```

4 matches found in sequence:
aatt49136 : Partial DNA clone AclI#3-134 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aatt49136 check: 6149 from: 1 to: 312

```

1      TGTGCGGTNARNGTGCGCTGCATGATTGCTTTGCACCGCTGTGACGCTTATCATCA
      |-----|-----|-----|
61      AGTCGTGCGCAGCTCCGTATATTCGCCGCGGATTCGCCGCGGACCGACCTGCTCGGA
      62 67      |-----|
      89 94
121     GCTGGCTTCGTGGGTATCGAATGCGCGGACCGCGGCTTTTCGAGATAGTTCAA
      145 150
181     CACCGGCAACGAGACGCCCTTG
      196 201

```

```

XX
XX MO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
PR
XX (UYVI-) UNIV VICTORIA.
PA
XX Nano FE;
PI
XX WPI; 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 48; 79pp: English.
XX
XX AAT49137 standard; DNA: 676 BP.
XX
XX AAT49137;
AC
XX
XX 10-NOV-1997 (first entry)
DT
XX Partial DNA clone AclI#3-166 encoding immunostimulatory peptide.
DE
XX Mycobacterium tuberculosis; immunostimulatory peptide: T cell epitope:
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS

```

14 matches found in sequence:
aatt49137 : Partial DNA clone AclI#3-166 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aatt49137 check: 1573 from: 1 to: 676

```

81      TNAATCCTATGATATATTGTCGGTAGCTTCCGCCCATGATTNGCTTTGCACCGCTGTTG
      |-----|-----|-----|
141     ACGCTTATCAATCAAGTCGTCGNCACCTCCGTATATTCGCCGTGNCATTTCTCCGCCGGG
      145 156 161
201     ACCGACTGCTC
      290 295
240     SCGACGCCACAGCGCGTTTTCGSAGATGTTCAACACCGGCAACGAGAGCCGCTTGG
      290 295
300     CGGCTCGGGGCTN

```

DT 10-NOV-1997 (first entry)
XX
DE Partial DNA clone Acil#3-166 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN W09700067-A1.
XX
PD 03-JAN-1997.
XX
PE 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI: 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
PS Claim 1; Page 49; 79pp: English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
CC
SQ Sequence 676 BP; 103 A; 191 C; 221 G; 112 T; 49 other:

AAT49137 Length: 676 March 5, 2002 14:19 Type: N Check: 1573 ..
Found using 'seq2-3' (pappu403.key)

1 AGGCCAATCGNTGATGCGACTCGAAGGGTTGCGCGCGATGACTGTTCCGAAGTTCA
|-----|
33 38

61 TCAGCACCCCTCGTTGGCGCAAGGGCAC

169 GKGKCGTGNCGGSCAKACANAGATTGGTTCACTGCGNATCGACACCNACGACCCGACG
|-----|
219 224

229 GTGCGCCCGCGAGACGAGCANACGNAATNGCAGCCGANTCNVACGNGGTGCGCCAGGA
|-----|
232 237 245 279 284

289 GGGACGNCCTACTTGTGTCGCGCGAGCCGACCGAGCTGTTNNCGGASGNCACGCGG
|-----|
303 308 342 347

349 AAGCGTTGACCCACTTGGTCAAGCAGCAGTGGNATCANGNATCANNGT

...
415 CCGTACGCGGTGTGCGGGGTGGGTCTTGTGGCACCGTGCACCTGTGGTGGCGCTTGGC
|-----|
465 470
475 GCGGTGTGGCGCTTGTGCGGGGTGGGTGGAGCCGCGGTGTGCGAGCGTGTCTGAAAT
|-----|
483 488 511 516 524
535 CGAGACCGGAGATCCATCGACAGCAAGTACTGCGNCTTGATGAGCGGCTCCGACTACCG
|-----|
533 558
595 GTGACTTCACTCCGACGCGAGACCGCATTCGCCGCGGSCGACAGAGGTGTGCCGACGTTGC
|-----|
615 620 621 626
655 ANASACGGTGGCGACAGCNSA

12 matches found in sequence:
aat49138 ; Partial DNA clone Acil#3-167 encoding immunostimulatory peptide.
(from "mycobactereng.seq")
TOIG of: aat49138. check: 5242 from: 1 to: 853

ID AAT49138 standard; DNA; 853 BP.
XX
XX AAT49138;
AC
XX
DT 10-NOV-1997 (first entry)
XX
XX Partial DNA clone Acil#3-167 encoding immunostimulatory peptide.
DE
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX W09700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
PE
XX
XX 15-JUN-1995; 95US-0000254.
PR
XX
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX Nano FE;
PI
XX
XX WPI: 1997-077347/07.
DR
XX
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
PS Claim 1; Page 50; 79pp: English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to

CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells. This
 CC sequence has homology to M. leprae DNA sequence within region B983.

XX Sequence 853 BP; 175 A; 247 C; 245 G; 145 T; 41 other:

AAAT49138 Length: 853 March 5, 2002 14:19 Type: N Check: 5242 ..
 Found using 'seq2-3' (pappu403.key)

```

67  CACCGAACCCNANMGATGCCGCGCGCCGCTGTAGAAAGGGAATCGCCAGTG      |-----|
127 CTGACTGCCATCATCCGACGCTTGAAKCKTTTGGGACGCTCCGTAGTGCAATC      |-----|
187 GTCGATTATGGGCTGTACTACTGCGAATTCGAGTCTGTGGTGTGCGSSGC      |-----|
316 CATTGGCAAGTTACTGCCGTGAGSSCAGCAGCAGGCGNCGACANGATGACGATGAC      |-----|
376 ATCGCCAGNCACTASAAAATCSCGTCGATGCTNCGCGCAACGTGATTCGANG      |-----|
493 TCCTCCGAGACAGACCATCAGCAANNGACCGCTTCCAGTGAGATCGGCGCGCTGGAC      |-----|
553 AANTCCSAATCNGCGGTTGGCCGCAATTNGCCCAAGGAGAGATCG      |-----|
627 GGTGGGTGGGCTGGGACCCGCGNNTTGCAACGGTTGTGATTCACACTCAAGCATGCTC      |-----|
687 GGTGACTTCAAAACCAACATTTGGCGGCAACGACATCATGCACTCCGGCGCAAT      |-----|
747 TTGACAGACCAAGTCAACACGGGTGATCAGATCGACGCTGGCGCGCAATTTGAACAT      |-----|
807 TGGCCGACAGACGCGNAGACCAAGGAKCA      |-----|

```

3 matches found in sequence:
 aat49139 ; Partial DNA clone AclI#3-204 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")

```

TOIG of: aat49139 check: 1026 from: 1 to: 209
ID  AAT49139 standard; DNA; 209 BP.
AC  AAT49139;
XX  10-NOV-1997 (first entry)
DT  Partial DNA clone AclI#3-204 encoding immunostimulatory peptide.
DE  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX  Mycobacterium tuberculosis.
OS  Mycobacterium tuberculosis.
XX  W09700067-A1.
XX  03-JAN-1997.
XX  14-JUN-1996; 96WO-US10375.
XX  15-JUN-1995; 95US-0000254.
XX  (UYVI-) UNIV VICTORIA.
XX  Nano FE;
XX  WPI: 1997-077347/07.
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
XX  useful in vaccines, diagnostic skin test, immunoassay and gene
XX  isolation
XX  Claim 1; Page 50; 79pp: English.
XX  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX  which encode partial sequences of immunostimulatory peptides. Each of
XX  the clones encode at least one immunostimulatory T cell epitope. The
XX  clones were identified by testing over 300 fusion clones (alkaline
XX  phosphatase-M. tuberculosis peptide fusions) for their ability to
XX  stimulate interferon (IFN)-gamma production. 80 clones were initially
XX  designated to have some ability to stimulate IFN-gamma production, of
XX  which 76 are shown in AAT49100-175. These sequences can be used to
XX  obtain the full length M. tuberculosis genes and corresponding proteins
XX  using standard techniques. The peptides are useful in vaccines, as
XX  reagents in an improved tuberculin skin test (especially using peptides
XX  different from those used in vaccines so as to allow differentiation
XX  between vaccinated and infected subjects) and as immunoassay reagents
XX  for detecting specific antibodies. An advantage of these peptides is
XX  that they stimulate production of IFN-gamma (critical for a protective
XX  immune response to M. tuberculosis) by CD4-positive T cells.
XX  Sequence 209 BP; 46 A; 72 C; 54 G; 32 T; 5 other:
XX  AAT49139 Length: 209 March 5, 2002 14:19 Type: N Check: 1026 ..
XX  Found using 'seq2-3' (pappu403.key)

```

```

19  GANNAATCAGCAGNDCCGCCACGTTGCAAGACGGCGCAATCTTGCCTGCTGCACAC      |-----|
79  GACTTCCGCCCACTTCAACATCTTGACATCGACGCGCABABGCCGAGGTCCGCGG      |-----|
139 ATGATGAAGACCACTTGCCTCCGCGCGCTTCAATCGACCGGACGCTCGACGCGTC      |-----|
140 189 194
199 GACACCAACTA

```

4 matches found in sequence:
aat49140 ; Partial DNA clone Acil#3-206 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49140 check: 7428 from: 1 to: 166

ID AAT49140 standard; DNA; 166 BP.
XX AAT49140;
AC
XX
XX
DT 26-SEP-1997 (first entry)
XX
DE Partial DNA clone Acil#3-206 encoding immunostimulatory peptide.
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
PN WO9700067-A1.
XX
XX 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE:
PI WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PI useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 51; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence has homology to M. leprae DNA sequence within chromosome region
CC MD0092.
XX
XX Sequence 166 BP; 23 A; 51 C; 56 G; 32 T; 4 other;
SQ

AAT49140 Length: 166 March 5, 2002 14:19 Type: N Check: 7428 ..
Found using 'seq2-3' (pappu403.key)

1
1 AGATGCTACGTACGACGAACCCGCCCAACCGCGCCCGAGCTGTTGTCAGGCGCTGA
3 8
61 AGNCNCTGCTGCAGCGAGCGCTGCGCCGCTGCGGAGTCGCGCTGCGCGCTGT
78 83 91 96

121 ACTTCAGCCGCGGATGTCGCCGCCGAGATGCTGATCATCGGTT
131 136

1 match found in sequence:
aat49141 ; Partial DNA clone Acil#3-214 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49141 check: 5999 from: 1 to: 221

ID AAT49141 standard; DNA; 221 BP.
XX AAT49141;
AC
XX
XX
DT 26-SEP-1997 (first entry)
XX
DE Partial DNA clone Acil#3-214 encoding immunostimulatory peptide.
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
PN WO9700067-A1.
XX
XX 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE:
PI WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PI useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 51; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX Sequence 221 BP; 47 A; 73 C; 37 G; 40 T; 24 other;
SQ

AAT49141 Length: 221 March 5, 2002 14:19 Type: N Check: 5999 ..
Found using 'seq2-3' (pappu403.key)

158 CTCTKCGCCGCGTGCACACGCGCATGTCGCGCGGACACCGCGATGTTGACGACCGCG
208 213
218 TGCA

2 matches found in sequence:

aat49142 : Partial DNA clone AcII#3-281 encoding immunostimulatory peptide.
(from "mycobacteriung.seq")
TOIG of: aat49142 check: 1669 from: 1 to: 117

```
ID      AAT49142 standard; DNA: 117 BP.
XX
XX
AC      AAT49142:
XX
XX      16-OCT-1997 (first entry)
XX
XX      Partial DNA clone AcII#3-281 encoding immunostimulatory peptide.
DE
XX      Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM      stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW      tuberculin skin test; ds.
XX
XX      Mycobacterium tuberculosis.
OS
XX      WO9700067-AI.
XX
XX      03-JAN-1997.
XX
XX      14-JUN-1996; 96MO-US10375.
XX
XX      15-JUN-1995; 95US-0000254.
XX
XX      (UYVI-) UNIV VICTORIA.
XX
XX      Nano FE;
XX
XX      WPI: 1997-077347/07.
DR
XX      New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT      useful in vaccines, diagnostic skin test, immunoassay and gene
PT      isolation
XX
XX
PS      Claim 1; Page 52; 79pp; English.
XX
XX      AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC      CC which encode partial sequences of immunostimulatory peptides. Each of
CC      the clones encode at least one immunostimulatory T cell epitope. The
CC      clones were identified by testing over 300 fusion clones (alkaline
CC      phosphatase-M. tuberculosis peptide fusions) for their ability to
CC      stimulate interferon (IFN)-gamma production. 80 clones were initially
CC      designated to have some ability to stimulate IFN-gamma production, of
CC      which 76 are shown in AAT49100-175. These sequences can be used to
CC      obtain the full length M. tuberculosis genes and corresponding proteins
CC      using standard techniques. The peptides are useful in vaccines, as
CC      reagents in an improved tuberculin skin test (especially using peptides
CC      different from those used in vaccines so as to allow differentiation
CC      between vaccinated and infected subjects) and as immunoassay reagents
CC      for detecting specific antibodies. An advantage of these peptides is
CC      that they stimulate production of IFN-gamma (critical for a protective
CC      immune response to M. tuberculosis) by CD4-positive T cells. This
CC      sequence is similar to sequence within M. leprae genomic clone B983.
XX
XX      Sequence 117 BP; 20 A; 24 C; 36 G; 16 T; 21 other;
SQ
AAT49142 length: 117 March 5, 2002 14:19 Type: N Check: 1669 ..
Found using 'seq2-3' (pepnu403.key)
```

```
1      CGGCGCGNCAAYYNGCGCHNCGGYGAGAGTCGNYAAGTCGCCAAGGTAACGCT
14      44 49 55 60
61      GATGCGATGGNACAGCAAGTATTGTGNACTTCACCGGTGHTGCTHGC
-----
```

4 matches found in sequence:

aat49143 : Partial DNA clone BsaHI#1-21 encoding immunostimulatory peptide.
(from "mycobacteriung.seq")
TOIG of: aat49143 check: 7190 from: 1 to: 385

```
ID      AAT49143 standard; DNA: 385 BP.
XX
XX
AC      AAT49143:
XX
XX      16-OCT-1997 (first entry)
XX
XX      Partial DNA clone BsaHI#1-21 encoding immunostimulatory peptide.
DE
XX      Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM      stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW      tuberculin skin test; ds.
XX
XX      Mycobacterium tuberculosis.
OS
XX      WO9700067-AI.
XX
XX      03-JAN-1997.
XX
XX      14-JUN-1996; 96MO-US10375.
XX
XX      15-JUN-1995; 95US-0000254.
XX
XX      (UYVI-) UNIV VICTORIA.
XX
XX      Nano FE;
XX
XX      WPI: 1997-077347/07.
DR
XX      New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT      useful in vaccines, diagnostic skin test, immunoassay and gene
PT      isolation
XX
XX
PS      Claim 1; Page 52; 79pp; English.
XX
XX      AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC      CC which encode partial sequences of immunostimulatory peptides. Each of
CC      the clones encode at least one immunostimulatory T cell epitope. The
CC      clones were identified by testing over 300 fusion clones (alkaline
CC      phosphatase-M. tuberculosis peptide fusions) for their ability to
CC      stimulate interferon (IFN)-gamma production. 80 clones were initially
CC      designated to have some ability to stimulate IFN-gamma production, of
CC      which 76 are shown in AAT49100-175. These sequences can be used to
CC      obtain the full length M. tuberculosis genes and corresponding proteins
CC      using standard techniques. The peptides are useful in vaccines, as
CC      reagents in an improved tuberculin skin test (especially using peptides
CC      different from those used in vaccines so as to allow differentiation
CC      between vaccinated and infected subjects) and as immunoassay reagents
CC      for detecting specific antibodies. An advantage of these peptides is
CC      that they stimulate production of IFN-gamma (critical for a protective
CC      immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX      Sequence 385 BP; 85 A; 103 C; 117 G; 80 T; 0 other;
SQ
AAT49143 length: 385 March 5, 2002 14:19 Type: N Check: 7190 ..
Found using 'seq2-3' (pepnu403.key)
```

```
71      ATGAGCGCAACGGGAATGATGTGAGTTACCTCATCTATTAGCACCCGCGCTGACA
121      121 126
131      CTTTATGCTTCCGGCTCGTGTAGTTGTGTGGAAATTGTGAGCGGATTAACAATTTCGACG
187
191      -|
      ACAGGAAACAGCTGTAGACATGATTTGACGAATTTGAATACAGCTCATATGAAATTTC
-----
```

192

251 GAGCTGGTACCGGGGATCTCTAGAGTCCTTCGCCCGGGGTCGCGCATCAGGGCCA
292 297

311 GTGCATGCCAAGCGGGGTACCGGGCCCTAGCTCTTCAGACATCGCGTGTGACCC
336 341

371 AGAGACCGGACGGG

2 matches found in sequence:
aat49144 ; Partial DNA clone HinpI#1-12 encoding immunostimulatory peptide.
(from "Mycobacterieng.seq")
TOIG of: aat49144 check: 6592 from: 1 to: 285

ID AAT49144 standard; DNA; 285 BP.

XX AAT49144;

DT 16-OCT-1997 (first entry)

DE Partial DNA clone HinpI#1-12 encoding immunostimulatory peptide.

XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;

XX Mycobacterium tuberculosis.

PN WO9700067-A1.

PD 03-JAN-1997.

PE 14-JUN-1996; 96WO-US10375.

PR 15-JUN-1995; 95US-0000254.

PA (UYVI-) UNIV VICTORIA.

XX Nano FE;

DR WPI; 1997-077347/07.

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation

PS Claim 1; Page 53; 79pp; English.

XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence encodes a protein that is possibly similar to tyrosine
CC phosphatase.

XX Sequence 285 BP; 47 A; 101 C; 87 G; 41 T; 9 other;

AAT49144 Length: 285 March 5, 2002 14:19 Type: N Check: 6592 ..
Found using 'seq2-3' (pappu403.key)

169 TCAGTACCCCGGGCAATACGCGCCGTATGNCAGTGGGTCAAGGGGTGCAACGTTGCT
219 224

229 TCGCGTATCGGCGGTGATCGCGGTATGCGCTGTTCATCGGCGCGTTCT
248 253

5 matches found in sequence:
aat49145 ; Partial DNA clone HinpI#1-142 encoding immunostimulatory peptide.
(from "Mycobacterieng.seq")
TOIG of: aat49145 check: 7023 from: 1 to: 186

ID AAT49145 standard; DNA; 186 BP.

XX AAT49145;

DT 16-OCT-1997 (first entry)

DE Partial DNA clone HinpI#1-142 encoding immunostimulatory peptide.

XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;

XX Mycobacterium tuberculosis.

PN WO9700067-A1.

PD 03-JAN-1997.

PE 14-JUN-1996; 96WO-US10375.

PR 15-JUN-1995; 95US-0000254.

PA (UYVI-) UNIV VICTORIA.

XX Nano FE;

DR WPI; 1997-077347/07.

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation

PS Claim 1; Page 53; 79pp; English.

XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence encodes a protein that is possibly similar to tyrosine
CC phosphatase.

XX Sequence 186 BP; 36 A; 53 C; 65 G; 25 T; 7 other;

AAT49145 Length: 186 March 5, 2002 14:19 Type: N Check: 7023 ..
Found using 'seq2-3' (pappu403.key)

...

12 TGGCGCCGGTGAACGTGAGCNCCTGNCATNAGCCCCAGCCGAGACGACGATGCCCGG |-----|
62 67

72 AGGAGTATGACAGACTGGGAGAGCCCGACGACTATGACGACTATGAGAGCGCG |-----|
92 97 107 112 116 121

132 CAGACCGAGAGCGCCGACGCGTCCGATCCTCGGACCGCGGTTGGCGGTTCGTT |-----|
162 167

6 matches found in sequence:
aat49146 : Partial DNA clone HinfI#1-144 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49146 check: 714 from: 1 to: 402

ID AAT49146 standard; DNA: 402 BP.

XX AAT49146:

XX 16-OCT-1997 (first entry)

DE Partial DNA clone HinfI#1-144 encoding immunostimulatory peptide.

XX

KM Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

XX

OS Mycobacterium tuberculosis.

XX

PN WO9700067-A1.

XX

PD 03-JAN-1997.

XX

PF 14-JUN-1996; 96WO-US10375.

XX

PR 15-JUN-1995; 95US-0000254.

XX

PA (UYVI-) UNIV VICTORIA.

XX

PI Nano FE;

XX

DR WPI; 1997-077347/07.

XX

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

XX

PT useful in vaccines, diagnostic skin test, immunoassay and gene

PT isolation

XX

PS Claim 1; Page 53-54; 79pp; English.

XX

CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,

CC which encode partial sequences of immunostimulatory peptides. Each of

CC the clones encode at least one immunostimulatory T cell epitope. The

CC clones were identified by testing over 300 fusion clones (alkaline

CC phosphatase-M. tuberculosis peptide fusions) for their ability to

CC stimulate interferon (IFN)-gamma production. 80 clones were initially

CC designated to have some ability to stimulate IFN-gamma production, of

CC which 76 are shown in AAT49100-175. These sequences can be used to

CC obtain the full length M. tuberculosis genes and corresponding proteins

CC using standard techniques. The peptides are useful in vaccines, as

CC reagents in an improved tuberculin skin test (especially using peptides

CC different from those used in vaccines so as to allow differentiation

CC between vaccinated and infected subjects) and as immunoassay reagents

CC for detecting specific antibodies. An advantage of these peptides is

CC that they stimulate production of IFN-gamma (critical for a protective

CC Immune response to M. tuberculosis) by CD4-positive T cells. This

CC

CC sequence has similarity to M. leprae DNA sequence within chromosome

CC

XX

SQ

Sequence 402 BP; 84 A; 124 C; 108 G; 69 T; 17 other;

...

139

TGGCCACATCTGGGTGATCGAGTGAGCAGGTGACCGCCTATTGTCACGCTTAGA

199

CCTGATCGCCGGCTTNCACRGASNGCGCGGCTGACGCCCTGCTGGGGAACATC

259

TCCGCTTTCGCCCGCAGGYCAAAACCTTCATSAAGCACAACGAACTGAACATGTC

319

CTCAGCNGCGCATCCTSACSCCTGTTG

...

5 matches found in sequence:

aat49147 : Partial DNA clone HinfI#1-200 encoding immunostimulatory peptide.

(from "mycobacterieng.seq")

TOIG of: aat49147 check: 2183 from: 1 to: 468

ID AAT49147 standard; DNA: 468 BP.

XX AAT49147:

XX 16-OCT-1997 (first entry)

DE Partial DNA clone HinfI#1-200 encoding immunostimulatory peptide.

XX

KM Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

XX

KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;

XX

KM tuberculin skin test; ds.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9700067-A1.

XX

PD 03-JAN-1997.

XX

PF 14-JUN-1996; 96WO-US10375.

XX

PR 15-JUN-1995; 95US-0000254.

XX

PA (UYVI-) UNIV VICTORIA.

XX

PI Nano FE;

XX

DR WPI; 1997-077347/07.

XX

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

XX

PT useful in vaccines, diagnostic skin test, immunoassay and gene

XX

PT isolation

XX

PS Claim 1; Page 54; 79pp; English.

XX

CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4 positive T cells.

XX Sequence 468 BP, 61 A; 126 C; 161 G; 103 T; 17 other:
 50
 AAT49147 Length: 468 March 5, 2002 14:19 Type: N Check: 2183 ..
 Found using 'seq2-3' (pappu403.key)

108 GTGTTCATGCTGGACACACCTGCGCBACGATCNSCGCTGAGCCGATCGGCTTACG
 158 163
 168 TCGGCTGGNNCCCTGTGTGGTTTCGACGCTCGCTGTTGGTGTCTGTGTGTCGAA
 228 GGTGTGCATCAGCTTCTGTGTGTTCGCTGCTGATTCGGTAAACGCTGGTACCGAGA
 240 245
 288 CGATGCCCCCGGCTGCGACTGCGGTTGTCTCTGTGCGNNCGCCGCGCATCTCGCGC
 291
 348 GCGNGTGTGACTGGGTACTGCTGCGAGCAGCAGCGACGACTTCATCGGGAATTCC
 374 379
 408 CGGGAAGTGGTGCATCTTGCCA

4 matches found in sequence:
 aat49148 : Partial DNA clone HinpI#2-23 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49148 check: 9946 from: 1 to: 417

ID AAT49148 standard; DNA: 417 BP.
 XX
 AC AAT49148:
 XX
 DT 16-OCT-1997 (first entry)
 XX
 DE Partial DNA clone HinpI#2-23 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9700067-A1.
 XX
 PD 03-JAN-1997.

XX
 PF 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE:
 XX
 DR WPI, 1997-077347/07.
 XX
 PT New immuno-stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1: Page 55; 79pp; English.

CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells. This
 CC sequence is similar to sequence in M. leprae genomic clone
 CC MD0009-0(B13).
 XX
 SO Sequence 417 BP; 89 A; 139 C; 110 G; 79 T; 0 other:
 AAT49148 Length: 417 March 5, 2002 14:19 Type: N Check: 9946 ..
 Found using 'seq2-3' (pappu403.key)

219 AATGGTTTCCCGTTGTCCACCGAGAAATCCCGCTGATTCAGGCCAATATACGCTCCCT
 269 274
 279 ATCCGAGTGTACACAGTGGGTGCAAGCGGAGTGGCCAACTGTGCTACCAACCCCG
 307 312
 339 ACTACGTTTTCACAACTCGACTCGACCGGTGAATCAAAACCGGGAGTGTGATGCCAG
 383 388
 399 CAACCTATGTGGGACTG

6 matches found in sequence:
 aat49149 : Partial DNA clone HinpI#2-143 encoding immunostimulatory peptide.
 (from "mycobacterieng.seq")
 TOIG of: aat49149 check: 5923 from: 1 to: 279

ID AAT49149 standard; DNA: 279 BP.
 XX
 AC AAT49149:
 XX
 DT 16-OCT-1997 (first entry)
 XX
 DE Partial DNA clone HinpI#2-143 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW


```

KW      tuberculin skin test; ds.
XX
OS      Mycobacterium tuberculosis.
XX
PN      WO9700067-A1.
XX
PD      03-JAN-1997.
XX
PF      14-JUN-1996; 96WO-US10375.
XX
PR      15-JUN-1995; 95US-0000254.
XX
PA      (UYVI-) UNIV VICTORIA.
XX
PI      Nano FE;
XX
DR      WPI; 1997-077347/07.
XX
PT      New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT      useful in vaccines, diagnostic skin test, immunoassay and gene
PT      isolation
XX
PS      Claim 1; Page 55; 79pp; English.
XX
CC      AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC      which encode partial sequences of immunostimulatory T cell epitope. Each of
CC      the clones encode at least one immunostimulatory T cell epitope. The
CC      clones were identified by testing over 300 fusion clones (alkaline
CC      phosphatase-M. tuberculosis peptide fusions) for their ability to
CC      stimulate interferon (IFN)-gamma production. 80 clones were initially
CC      designated to have some ability to stimulate IFN-gamma production, of
CC      which 76 are shown in AAT49100-175. These sequences can be used to
CC      obtain the full length M. tuberculosis genes and corresponding proteins
CC      using standard techniques. The peptides are useful in vaccines, as
CC      reagents in an improved tuberculin skin test (especially using peptides
CC      different from those used in vaccines so as to allow differentiation
CC      between vaccinated and infected subjects) and as immunoassay reagents
CC      for detecting specific antibodies. An advantage of these peptides is
CC      that they stimulate production of IFN-gamma (critical for a protective
CC      immune response to M. tuberculosis) by CD4-positive T cells. This
CC      sequence is similar to M. leprae genomic clone B1529.
XX
SQ      Sequence 279 BP; 52 A; 92 C; 82 G; 53 T; 0 other;
XX
AAT49149 Length: 279 March 5, 2002 14:19 Type: N Check: 5923
Found using 'seg2-3' (pappu403.key)
1      CGGTGAGCCGATGACGCTCTGACATTACCGCAACCAACGCTCAGCGTGCTCCCTTGA
15      20
121     TGCACATGTGAGAGCCCGCCAGCCAGCAGTAGCGGCGGCTAATCCGGCGGTACACT
181     TCGTGGCAATGCGCACCGCGCGGAGCTCGGGGCGATGACAGCTTTGTCTCGAAGTACA
189     194      204      209      217
241     ACCTGAATTTACCAACACCTCAATGACGCCGCTGCTGTGA
264     269
-----
3 matches found in sequence:
aat49150 : Partial DNA clone HinpI#2-145 encoding immunostimulatory peptide.
(from "mycobacter.res")
TOIG of: aat49150 check: 9554 from: 1 to: 324
ID      AAT49150 standard; DNA; 324 BP.
XX

```

```

AC      AAT49150;
XX
DT      16-OCT-1997 (first entry)
XX
DE      Partial DNA clone HinpI#2-145 encoding immunostimulatory peptide.
XX
KW      Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW      stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX      tuberculin skin test; ds.
XX
OS      Mycobacterium tuberculosis.
XX
PN      WO9700067-A1.
XX
PD      03-JAN-1997.
XX
PF      14-JUN-1996; 96WO-US10375.
XX
PR      15-JUN-1995; 95US-0000254.
XX
PA      (UYVI-) UNIV VICTORIA.
XX
PI      Nano FE;
XX
DR      WPI; 1997-077347/07.
XX
PT      New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT      useful in vaccines, diagnostic skin test, immunoassay and gene
PT      isolation
XX
PS      Claim 1; Page 56; 79pp; English.
XX
CC      AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC      which encode partial sequences of immunostimulatory T cell epitope. Each of
CC      the clones encode at least one immunostimulatory T cell epitope. The
CC      clones were identified by testing over 300 fusion clones (alkaline
CC      phosphatase-M. tuberculosis peptide fusions) for their ability to
CC      stimulate interferon (IFN)-gamma production. 80 clones were initially
CC      designated to have some ability to stimulate IFN-gamma production, of
CC      which 76 are shown in AAT49100-175. These sequences can be used to
CC      obtain the full length M. tuberculosis genes and corresponding proteins
CC      using standard techniques. The peptides are useful in vaccines, as
CC      reagents in an improved tuberculin skin test (especially using peptides
CC      different from those used in vaccines so as to allow differentiation
CC      between vaccinated and infected subjects) and as immunoassay reagents
CC      for detecting specific antibodies. An advantage of these peptides is
CC      that they stimulate production of IFN-gamma (critical for a protective
CC      immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ      Sequence 324 BP; 33 A; 98 C; 124 G; 63 T; 6 other;
XX
AAT49150 Length: 324 March 5, 2002 14:19 Type: N Check: 9554
Found using 'seg2-3' (pappu403.key)
1      CGGCGCGGCGCGCTGCTGTAAGCTTGAGAAATGGGTGAGCGAGCTGCCACACACAG
10      15
61      GGACC
153     CGGCGGTGTTGGGCTTCCGATTGCTGCGCTGGGTGCTGACCAACCGCGGACACACAG
203     208
213     GTGGGTGGGCTGGGCTACGGCTGCTATTCTGGGCTGGTGTCTACG
265     TGTTCGCTGAGATCGCGAGCTGTGTCNCGGCGGCTGTGTGGCATGNCGACGACGTGC
315     320
-----
1-----1

```

325

4 matches found in sequence:
aat49152 : Partial DNA clone HinfI#3-28 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49152 check: 5460 from: 1 to: 293

ID AAT49152 standard; DNA; 293 BP.
XX
XX AAT49152;
XX
DT 16-OCT-1997 (first entry)
XX
DE Partial DNA clone HinfI#3-28 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; Immunostimulatory peptide; T cell epitope;
RV stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
XX W09700067-A1.
XX
XX 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
PI
XX WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 57; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence is similar to M. leprae sequence in genomic clones MD0085 and gp
CC AAU00013 cds 27 of B1496 region.
XX
XX
SQ Sequence 293 BP; 68 A; 104 C; 73 G; 48 T; 0 other:
AAT49152 Length: 293 March 5, 2002 14:19 Type: N Check: 5460 ..
Found using 'seq2-3' (pappu403.key)

1 CCACACACACAATCTACGTCGTAATGCACTCCTAAGTCCATCCGCTGATGCGAAG
46 51
61 GACAGCACCGAGCGCCACGCGATATACATCGTCGCGCGGTGCACAAAGCACATCAT

121 CATGAGCTGCTCCACTACGCGCTAC 90 95
...
174 CGATGTCGACTGCGCACCCAGCATGTCCTACAGCGGTGCTCTGTCGACACGCGGTGG
224 229
234 TCGGTGGGGGCTCAGGGGCCACACCAACACCATGCGTGCCTGACGTGAGCCCGAGAC
279 284
294

6 matches found in sequence:
aat49153 : Partial DNA clone HinfI#3-30 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49153 check: 3023 from: 1 to: 815

ID AAT49153 standard; DNA; 815 BP.
XX
XX AAT49153;
XX
DT 16-OCT-1997 (first entry)
XX
DE Partial DNA clone HinfI#3-30 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; Immunostimulatory peptide; T cell epitope;
RV stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX W09700067-A1.
XX
XX 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
PI
XX WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 57-58; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence is similar to M. leprae chromosome sequence in B963 region.

```

XX      SQ      Sequence 815 BP; 127 A; 264 C; 245 G; 125 T; 54 other;
          AAT49153 Length: 815 March 5, 2002 14:19 Type: N Check: 3023
          Found using 'seq2-3' (pappu403.key)

...

20      CGTGGCGGAGGCTACAYCCAGACANTCCGGAGCCGGCGTCYCGCTGGAGACCCGTGC
          |-----|
          70 75

80      TCCTACAYACCGGCGRGGGGGCTTGCCACGSCCGGACACCCCACTACCCNCGCGGG
          |-----|
          138

140     CGCCACCRRTTGGCCCGTTNMGTGACCCGAGNCTTCCCGGACCGCTCGATGTCCAGCCG
          143
          199

200     -----|
          TGCCCCCTTAATCCCGATGGGCGCGMCGCGACKCCGGGCATCCTAAGTCTGGGC
          204

376     TTGGAATCCGSAAGCTGCKACTNGYCGAMACTGTCTGCAAGGTAAGTGTGCGCG
          |-----|
          426 431

436     GGTGGTGGTGCTTGGCGGTGTRGCGNCTGNCRGCCGGCGG

482     CRGCTTACCGGAAGTACTACTAATACCGTGTCCGRTATTTTCTSTAGGCGCTGCC
          |-----|
          532 537

542     CTGTACCCAGAGASAAAGTCCAGATCATGTGCTGCGGGTCTCTATCGACAAGATC
          |-----|
          611 616

602     GAGCCGCGCGGCGACAAGATGCGAGTCACGTTGCACTANCAAGCAAAATACAGGTGCC
          662     GCCCA

...

-----
1 match found in sequence:
aat49154 : Partial DNA clone HinfI#3-34 encoding immunostimulatory peptide.
(from "Mycobacterieng.seq")
T0IG of: aat49154 check: 4691 from: 1 to: 117

ID      AAT49154 standard; DNA: 117 BP.
XX
XX      AC      AAT49154;
XX
XX      DT      16-OCT-1997 (first entry)
XX
XX      DE      Partial DNA clone HinfI#3-34 encoding immunostimulatory peptide.
XX
XX      KW      Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
XX      KW      stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX      KW      tuberculin skin test; ds.
XX
XX      OS      Mycobacterium tuberculosis.
XX
XX      PN      WO9700067-A1.
XX
XX      PD      03-JAN-1997.
XX
XX      PF      14-JUN-1996; 96WO-US10375.
XX
XX      PR      15-JUN-1995; 95US-0000254.

```

```

XX      PA      (UUVI-) UNIV VICTORIA.
XX      PI      Nano FE;
XX      DR      MPI; 1997-077347/07.
XX
XX      PT      New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
XX      PT      useful in vaccines, diagnostic skin test, immunoassay and gene
XX      PT      isolation
XX
XX      PS      Claim 1; Page 58; 79pp; English.
XX
XX      CC      AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX      CC      which encode partial sequences of immunostimulatory peptides. Each of
XX      CC      the clones encode at least one immunostimulatory T cell epitope. The
XX      CC      clones were identified by testing over 300 fusion clones (alkaline
XX      CC      phosphatase-M. tuberculosis peptide fusions) for their ability to
XX      CC      stimulate interferon (IFN)-gamma production. 80 clones were initially
XX      CC      designated to have some ability to stimulate IFN-gamma production, of
XX      CC      which 76 are shown in AAT49100-175. These sequences can be used to
XX      CC      obtain the full length M. tuberculosis genes and corresponding proteins
XX      CC      using standard techniques. The peptides are useful in vaccines, as
XX      CC      reagents in an improved tuberculin skin test (especially using peptides
XX      CC      different from those used in vaccines so as to allow differentiation
XX      CC      between vaccinated and infected subjects) and as immunoassay reagents
XX      CC      for detecting specific antibodies. An advantage of these peptides is
XX      CC      that they stimulate production of IFN-gamma (critical for a protective
XX      CC      immune response to M. tuberculosis) by CD4-positive T cells. This
XX      CC      sequence is similar to sequence in M. leprae genomic clone B2168
XX      CC      (AAU0018 cds 9).
XX
XX      SQ      Sequence 117 BP; 20 A; 47 C; 31 G; 19 T; 0 other;
          AAT49154 Length: 117 March 5, 2002 14:19 Type: N Check: 4691
          Found using 'seq2-3' (pappu403.key)

...

9      TCGTGCGCGCGACATGACATATCAGCCGCGCCGCACTGCTGACCTGATGCCACACA
          |-----|
          59 64

69      GCTGAGGCGCCCTACCGGCTGCAATTCAATTCACCCGCTGGGCTCG

...

-----
1 match found in sequence:
aat49155 : Partial DNA clone HinfI#3-41 encoding immunostimulatory peptide.
(from "Mycobacterieng.seq")
T0IG of: aat49155 check: 2574 from: 1 to: 242

ID      AAT49155 standard; DNA: 242 BP.
XX
XX      AC      AAT49155;
XX
XX      DT      16-OCT-1997 (first entry)
XX
XX      DE      Partial DNA clone HinfI#3-41 encoding immunostimulatory peptide.
XX
XX      KW      Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
XX      KW      stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX      KW      tuberculin skin test; ds.
XX
XX      OS      Mycobacterium tuberculosis.
XX
XX      PN      WO9700067-A1.
XX
XX      PD      03-JAN-1997.
XX
XX      PF      14-JUN-1996; 96WO-US10375.
XX
XX

```

PR 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
PI
XX WPI: 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1: Page 58; 79pp; English.
PS
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence encodes a protein similar to antigen 85 complex protein
CC subunit.
XX
SQ Sequence 242 BP; 47 A; 87 C; 63 G; 45 T; 0 other;
AAT4915 Length: 242 March 5, 2002 14:19 Type: N Check: 2574
Found using 'seq2-3' (pappu403.key)
1 AGGTGCTGCTTCATGCTGCGCCCAATCATCTTCTACACCGACTGTATCACCTT
21 26
61 CGCAGACAAACGGCCA
1 match found in sequence:
aat49156; Partial DNA clone HpaII#1-3 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49156 check: 8666 from: 1 to: 340
ID AAT49156 standard; DNA: 340 BP.
XX
XX AAT49156;
AC
XX
XX 16-OCT-1997 (first entry)
DT
XX
XX Partial DNA clone HpaII#1-3 encoding immunostimulatory peptide.
DE
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
PF
XX
XX 15-JUN-1995; 95US-0000254.
PR

XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
PI
XX WPI: 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1: Page 59; 79pp; English.
PS
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence encodes a protein similar to cytochrome C oxidase subunit II. It
CC is also similar to M. tuberculosis sequence clone MTCY 190.11C.
XX
SQ Sequence 340 BP; 83 A; 87 C; 101 G; 67 T; 2 other;
AAT49156 Length: 340 March 5, 2002 14:19 Type: N Check: 8666
Found using 'seq2-3' (pappu403.key)
193 CGCGGCTCAACACCGAGACCGGACCTACCTGAATTTGCAAGATGCGAGCGTGGGC
243 248
253 AGCAGACCGAATTCGGTGTCTGTGCTGCGCGCAACGCTA
1 match found in sequence:
aat49157; Partial DNA clone HpaII#1-8 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49157 check: 2098 from: 1 to: 262
ID AAT49157 standard; DNA: 262 BP.
XX
XX AAT49157;
AC
XX
XX 16-OCT-1997 (first entry)
DT
XX
XX Partial DNA clone HpaII#1-8 encoding immunostimulatory peptide.
DE
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
PF
XX

```
PR 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX Nano FE:
PI
XX
XX WPI: 1997-077347/07.
DR
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 59; 79pp; English.
PS
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX Sequence 262 BP; 60 A; 71 C; 73 G; 45 T; 13 other;
SQ
AAT49157 Length: 262 March 5, 2002 14:19 Type: N Check: 2098 ..
Found using 'seq2-3' (pappu403.key)
...
131 CGGTCCTCAAGCGCTGATGCAGGTCTTCCCGGACAGGTAAGCCTAAGACGACATTG |-----|
181 186
191 GCGCGAGCTTGCGGAGCAGCAGCTGCGGTATCCGAGAGACTGTTCAA
...
-----
1 match found in sequence:
aat49158 ; Partial DNA clone HpaII#1-10 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49158 check: 6864 from: 1 to: 241
ID AAT49158 standard; DNA: 241 BP.
XX
XX AAT49158;
AC
XX
XX 16-OCT-1997 (first entry)
DT
XX
XX Partial DNA clone HpaII#1-10 encoding immunostimulatory peptide.
DE
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
PF
XX
XX 15-JUN-1995; 95US-0000254.
PR
```

```
XX
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX Nano FE:
PI
XX
XX WPI: 1997-077347/07.
DR
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX Claim 1; Page 60; 79pp; English.
PS
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence encodes a protein similar to immunogenic proteins MPB64/MP764.
XX
XX Sequence 241 BP; 44 A; 92 C; 54 G; 40 T; 11 other;
SQ
AAT49158 Length: 241 March 5, 2002 14:19 Type: N Check: 6864 ..
Found using 'seq2-3' (pappu403.key)
...
149 TCGACAGCTGTTCGTGTCGCCGCGCACGACCTGACAGCATCTACCATCGTTCAGC |-----|
199 204
209 GCGAGCTGCGCAGCTCAGACCGGTTTCGTCGCG
...
-----
3 matches found in sequence:
aat49159 ; Partial DNA clone HpaII#1-13 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49159 check: 1735 from: 1 to: 243
ID AAT49159 standard; DNA: 243 BP.
XX
XX AAT49159;
AC
XX
XX 16-OCT-1997 (first entry)
DT
XX
XX Partial DNA clone HpaII#1-13 encoding immunostimulatory peptide.
DE
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
PF
XX
XX 15-JUN-1995; 95US-0000254.
PR
XX
XX (UYVI-) UNIV VICTORIA.
PA
```

```
XX Nano FE;
PI
XX WPI: 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PI isolation
XX
XX Claim 1: Page 60; 79pp; English.
PS
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory T cell epitope. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
CC
XX Sequence 243 BP; 30 A; 83 C; 78 G; 37 T; 15 other;
SQ
AAT49159 Length: 243 March 5, 2002 14:19 Type: N Check: 1735 ..
Found using 'seq2-3' (pappu403.key)
...
57 NTNCCGCCGNCAGNCAGNCTNCGATGTCCNGTATACCTGCGCGATCGCGCGG
|-----|
117 GGCTGCCGACACACGCTGNGCCGCCGCTCTCCGCCAATTCTGGTGNGCGCATN
|-----|
177 CCGGACAGCGCCGCCGAGCAGTGTGATGATGCGGTGCGGCGGCGGTGGC
182 187 208 213
237 TGCTGGC
-----
32 matches found in sequence:
aat49160: Partial DNA clone AcII#2-825 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49160 check: 6333 from: 1 to: 2346
ID AAT49160 standard; DNA: 2346 BP.
XX
XX AAT49160;
AC
XX
XX 16-OCF-1997 (first entry)
DT
XX
XX Partial DNA clone AcII#2-825 encoding immunostimulatory peptide.
DE
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 1623
FT /*tag= a
FT /note= "represented by "-" in the specification"
XX
XX PN MO9700067-A1.
XX
XX 03-JAN-1997.
PD
XX 14-JUN-1996; 96WO-US10375.
PE
XX 15-JUN-1995; 95US-0000254.
PR
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX Nano FE;
PI
XX WPI: 1997-077347/07.
DR
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PI isolation
XX
XX Claim 1: Page 61-62; 79pp; English.
PS
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory T cell epitope. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells. This
CC sequence is similar to M. tuberculosis sequence MTCY31.03c.
CC
XX Sequence 2346 BP; 333 A; 749 C; 797 G; 429 T; 38 other;
SQ
AAT49160 Length: 2346 March 5, 2002 14:19 Type: N Check: 6333 ..
Found using 'seq2-3' (pappu403.key)
...
9 ATTCGNACTTGGAGCAGCGCTTGGCGTGTGCTGATCATGANCATGACGCGCGCACCG
|-----|
69 GCAGCTTCCGCTCATGGGTGCTATCCCTGTGGCGCTGGCCATGAT
59 64
...
186 TGTCGCCGGGTCACTCAACGGCTGACCGTGTTCGGCTGCTCGGGCGCACACATTCGCTGTG
|-----|
246 GCGGATTTGGCGCGGAGTGTGAATTGCTGTGACCAACCATGTTCCAGCTGCCGGCGCGT
236 241
306 TGTTCTGTCGTCGTCGCGGATCACCATCCTNNNTNNGCTTGGCTGTGATNCNCAATTCGGCGCT
|-----|
311 316
314 319
366 GGGT
...
382 CGGTGAGTCCCGGCGCACATTGAGCTACCAACGGGATAGGNCAGACTACGAGCGACNGCT
|-----|
432 437
442 GCGCGAGAGATTCAGAACCTCGGCGGAACACTCCGACAAACGCTT
```

```

...
513  TGGGGTAACAGACATCAAGTATGATGCGCTTCTGTTCTTATCCGGGGTGTGTC      |-----| |---
563  568

573  --|
575  GCCAAGCGCAGCAAGCCACAGGGTGGTGCATTTGGGCATGCTGGGCGCTGAT      |-----| |---
720  GTGCTAGTTACCGTTGTTAGCCATCGCGCGCGGTGGCCGACGCTGGCAGGACAGCN      |-----| |---
770  775

780  ATTGCCACCCCTGATCAGGCGAGGGTCCAGTGCATTTGCTAAAGCCTGCTGAGAGCCTCG      |-----| |---
831  836

840  TTGCAGCAGACCTGCCGCCAGAGTGGCGGCATCGGGGTTTGGGCGTTCCGAGTGCAC      |-----| |---
883  888

900  CTTACAGTGGCTGGGTGCTGGGCGGCGCGGTGGGCGCTG      |-----| |---
1021  TCCGCGCGATTCGCTGATCCCTGGCCCTGGCGGTAATCGGCCCTGATGCGCGAGCAAG      |-----| |---
1026  1031

1081  AAMCAACCCGTCGTGTGGGGCGGCGCGGCGGTAAGAGCGGGTGTCCGACGCTGCG      |-----| |---
1106  1111

1141  GGTGATCCCTGGTATTCCTGCTCGGTGGCGGCGCGGTGCAGTGGCTGCTAGTACG      |-----| |---
1229  1234

1201  CGGACACGCTCCGCGAGCACCCGAGATCAGCGCTTACTGCGACGGGACCTGACCCGCGT      |-----| |---
1295  1300

1261  GGGGCCCTATTTGTACTGCAACGTGTGACACTGACGACTGCAGACCCCGCAGNGCGCA      |-----| |---
1341  1346

1321  GGGCGAATTGCCCGTAAGCGACGCTATCCCTGCGACGCTCTCGGTACCCGGAAGTCATTTC      |-----| |---
1384  1389  1397

1381  CCGGGCGCGCGGTGGTGGTGGAGTATACAGAGACCCCGCAACACACACACACCTT      |-----| |---
1441  GTTTCGG

1478  ACTGTGACCCCGCAGCGCGGCGGCTGACCGGGATTGTCTGCGAGTTGCTGACGTTGGTG      |-----| |---
1528  1533

1538  GTCGACCACTGGGTGAATACGGAGCTACGACGCGGGAATGGTC

1588  CGGCTATCTTTTACAGAGCGCGGCTCGACGNCNCCTTAAAGCGCGGTGCGGCCAAGC      |-----| |---

```

```

1648  GTCCGAAGAGCCGCCGACACCCGGGGACATCGCGGCAATCATGAA      1638  1643
...
1738  CAGGCGCGGGTGGAGGCTGCGGGCCAGGCGCGGACAGGCTGGCGGGGTAGAGCGCGTCG      |-----| |---
1788  1793  1794

-|
1798  CTGGCGGTTGTTGGTTGATGCGGTGACACACACCGACCGGGCTGCGCGCTCGAGAGCGG      |-----| |---
1843  1848

1858  GTGCTGAGATGATCGACCCCGCGGCTTGTGCGTTGCATCGCCGACGCGCATGTCGCG      |-----| |---
1897  1902  1909  1914  1912  1917

1918  GGCGCGCCAGAGATCGAGAGACGAGCCCGGCGGTGGTGTGCGTGGCGCTCGGCTTGGCC      |-----| |---
1962  1967

1978  GCCGAGACATTCAGACTGAGACTTGTGCTGACCGGCTCGG

...
2122  AACACGACCTGCCGGGACACCGCNGTGTGGGGCGGCGGNTGTAGCGCGCGCGCGCGG      |-----| |---
2172  2177

2182  GGGGACACCCGGSTGKCCGCGATCAGACGTGCTACCTCCGGMGTGTCGCGCGTGGCG      |-----| |---
2183  2188  2228  2233

2242  CCTGSCGGGATGCGCGGTGTCGCGGTGTCGACAGGGTTGACGCGCATGCCATACCC      |-----| |---
2306  2311

2302  ATACATGCTACCGGMCAGGAGCATATGTRKACCGAGCTCGG

-----
9 matches found in sequence:
aat49161; Partial DNA clone AclI#1-435 encoding immunostimulatory peptide.
(from "mycobactering.seq")
TOIG of: aat49161 check: 9768 from: 1 to: 841

ID  AAT49161 standard; DNA; 841 BP.
XX
AC  AAT49161;
XX
XX
DE  16-OCT-1997 (first entry)
XX
XX  Partial DNA clone AclI#1-435 encoding immunostimulatory peptide.
XX
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
OS
XX  WO9700067-A1.
XX  PN
XX  PD  03-JAN-1997.
XX  PF  14-JUN-1996; 96MO-US10375.
XX  PR  15-JUN-1995; 95US-0000254.

```

```

...
632 CGGACGCGCAACNCNCCTGTGGCGCAATCTGGGTNTTNTTCGCTGGCGCTGGCGCCGCGC |-----|
      (GGGACGCGCAACNCNCCTGTGGCGCAATCTGGGTNTTNTTCGCTGGCGCTGGCGCCGCGC 682 687
692 GNANGCNCGACGTGTGGCGCGCATTTGGCCCTATGACCCGAACCTGTAGCCCTGAGACCAATGGT |-----|
      (GNANGCNCGACGTGTGGCGCGCATTTGGCCCTATGACCCGAACCTGTAGCCCTGAGACCAATGGT 715 720
752 TCGCCTTCACCCCGCCCGCGGTTCCGGTGCTGGAATCTCTGACGCGTTGACGCGTGACG 1-
      (TCGCCTTCACCCCGCCCGCGGTTCCGGTGCTGGAATCTCTGACGCGTTGACGCGTGACG 810
812 CGGACCGCGGTGTGATGACATCGCGACCG
      (CGGACCGCGGTGTGATGACATCGCGACCG 815
-----
9 matches found in sequence:
aat49162 : Partial DNA clone AcII#1-2/23/9 encoding immunostimulatory peptide.
      (from "Mycobacterium.seq")
TOIG of: aat49162 Check: 4468 from: 1 to: 471
ID AAT49162 standard; DNA; 471 BP.
XX
XX AAT49162;
AC
XX
XX 16-OCT-1997 (first entry)
DT
XX
XX Partial DNA clone AcII#1-2/23/9 encoding immunostimulatory peptide.
DE
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
XX stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM tuberculin skin test; ds.
KW
XX
XX Mycobacterium tuberculosis.
OS
XX
XX WO9700067-A1.
PN
XX
XX 03-JAN-1997.
PD
XX
XX 14-JUN-1996; 96WO-US10375.
PR
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX Nano FE;
PI
XX
XX WPI; 1997-077347/07.
DR
XX
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
XX
XX Claim 1; Page 63; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is

```


CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ Sequence 471 BP; 85 A; 161 C; 146 G; 79 T; 0 other;

AAT49162 Length: 471 March 5, 2002 14:19 Type: N Check: 4468 ..
Found using 'seq2-3' (pappu403.key)

12 TCGGCTGACCGGAGTGATGATCACCACCTCAACGTGTGTGGCCTGCTGCGCTCACA
62 67

72 CGATCGGTGGACCAAGCGGCTGATCAGCGTATTCACCGGCTCGCGCAACG
93 98 105 110

132 CAAGACCGACATCTCCACAGCGCGTACACCAACGCGCGCGCTGCGCGCAT
148 153 183 188

192 CTCTGTGCAAGCTCGCGCGTGGCGAAGTGTTCGCGAGACGA

264 GACACGACTACTCTCGACAATCTGCTCACACGCTCGCGGACAATACAGGCGTGTGTC
314 319

324 CGCCAGGCTATGTAGCGCGCTTCTGCGCTTCTACTGTGACGTCGCTCAAGTTC
339 344 366 371

384 AACGCAAGGCGCGCCAGCGGTGTACATCAAGCTGCGC

3 matches found in sequence:
aat49163 ; Partial DNA clone Acl1#1-229/264 encoding immunostimulatory peptide
(from "Mycobacterin.seq")
TOIG of: aat49163 check: 3614 from: 1 to: 485

ID AAT49163 standard; DNA: 485 BP.

AC AAT49163;

DT 16-OCT-1997 (first entry)

DE Partial DNA clone Acl1#1-229/264 encoding immunostimulatory peptide.

KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KM stimulate; interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

OS Mycobacterium tuberculosis.

PN WO9700067-A1.

PD 03-JAN-1997.

PF 14-JUN-1996; 96WO-US10375.

PR 15-JUN-1995; 95US-0000254.

PA (UYVI-) UNIV VICTORIA.

PI Nano FE;
XX
DR WPI; 1997-077347/07.

XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation

XX Claim 1; Page 63-64; 79pp; English.

XX
PS
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.

XX Sequence 485 BP; 92 A; 114 C; 142 G; 111 T; 26 other;

AAT49163 Length: 485 March 5, 2002 14:19 Type: N Check: 3614 ..
Found using 'seq2-3' (pappu403.key)

1 KGTCTGCGNCCTTNCATCGCGTCCGNNRGGTNACTGTGCTGTGATCGGTCCGA
23 28

61 NGTATNANCAATGCGCA

291 TTGCAGATGCGCGCCAGCGCATCAGCTATGACGATTCGGGGCCACAAGTCCGC
341 346

351 GTTCTGCGGGTTTCTCTGACTACCTCAGATCAAGTGTGATTTTGATGATCGCGCC

411 AGTTATCGGGCCCATTCGACGATGATGACGTGACGACATGTTCTTTTGGCNATN
443 448

471 CGGGTGGCAATAGCG

3 matches found in sequence:
aat49164 ; Partial DNA clone Acl1#1-264 encoding immunostimulatory peptide.
(from "Mycobacterin.seq")
TOIG of: aat49164 check: 1123 from: 1 to: 469

ID AAT49164 standard; DNA: 469 BP.

AC AAT49164;

DT 16-OCT-1997 (first entry)

DE Partial DNA clone Acl1#1-264 encoding immunostimulatory peptide.

KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KM stimulate; interferon-gamma; IFN-gamma; production; vaccine;

OS Mycobacterium tuberculosis.

PN WO9700067-A1.

PD 03-JAN-1997.

```

PF 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
XX
XX WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
XX useful in vaccines, diagnostic skin test, immunoassay and gene
XX isolation
XX
XX Claim 1; Page 64; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX which encode partial sequences of immunostimulatory peptides. Each of
XX the clones encode at least one immunostimulatory T cell epitope. The
XX clones were identified by testing over 300 fusion clones (alkaline
XX phosphatase-M. tuberculosis peptide fusions) for their ability to
XX stimulate interferon (IFN)-gamma production. 80 clones were initially
XX designated to have some ability to stimulate IFN-gamma production, of
XX which 76 are shown in AAT49100-175. These sequences can be used to
XX obtain the full length M. tuberculosis genes and corresponding proteins
XX using standard techniques. The peptides are useful in vaccines, as
XX reagents in an improved tuberculin skin test (especially using peptides
XX different from those used in vaccines so as to allow differentiation
XX between vaccinated and infected subjects) and as immunoassay reagents
XX for detecting specific antibodies. An advantage of these peptides is
XX that they stimulate production of IFN-gamma (critical for a protective
XX immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX Sequence 469 BP; 96 A; 130 C; 141 G; 97 T; 5 other;
XX
AAT49164 Length: 469 March 5, 2002 14:19 Type: N Check: 1123 ..
Found using 'seq2-3' (pappu403.key)
1 GGCAGGCTCACTGAAGCCGAGAGCGGAAGAGCGCCCAATACGAAACCCCTCTCC
14 39
61 CGCGCGTTGGCCGATTCATTAAATGCAGC
...
270 ATACGACTCACTAATAGGAATTCGAGCTCGGTACCGGGATCCTTAGAGTCGCTTGG
320 325
330 TTGGCGGCGACGACGAGTCCAGCGGTGGCCGCCGCCGCGGCTCATACACCGCGCG
335 340
390 G
...
4 matches found in sequence:
aat49165; Partial DNA clone AclI#1-264C encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49165 check: 2477 from: 1 to: 290
ID AAT49165 standard; DNA: 290 BP.
XX
XX AAT49165;
AC
XX 16-OCT-1997 (first entry)
DT
XX
DE Partial DNA clone AclI#1-264C encoding immunostimulatory peptide.
XX
XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW

```

```

KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9700067-A1.
XX
XX 03-JAN-1997.
XX
XX 14-JUN-1996; 96WO-US10375.
XX
XX 15-JUN-1995; 95US-0000254.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX Nano FE;
XX
XX WPI; 1997-077347/07.
XX
XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
XX useful in vaccines, diagnostic skin test, immunoassay and gene
XX isolation
XX
XX Claim 1; Page 65; 79pp; English.
XX
XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
XX which encode partial sequences of immunostimulatory peptides. Each of
XX the clones encode at least one immunostimulatory T cell epitope. The
XX clones were identified by testing over 300 fusion clones (alkaline
XX phosphatase-M. tuberculosis peptide fusions) for their ability to
XX stimulate interferon (IFN)-gamma production. 80 clones were initially
XX designated to have some ability to stimulate IFN-gamma production, of
XX which 76 are shown in AAT49100-175. These sequences can be used to
XX obtain the full length M. tuberculosis genes and corresponding proteins
XX using standard techniques. The peptides are useful in vaccines, as
XX reagents in an improved tuberculin skin test (especially using peptides
XX different from those used in vaccines so as to allow differentiation
XX between vaccinated and infected subjects) and as immunoassay reagents
XX for detecting specific antibodies. An advantage of these peptides is
XX that they stimulate production of IFN-gamma (critical for a protective
XX immune response to M. tuberculosis) by CD4-positive T cells.
XX
XX Sequence 290 BP; 78 A; 77 C; 82 G; 52 T; 1 other;
XX
AAT49165 Length: 290 March 5, 2002 14:19 Type: N Check: 2477 ..
Found using 'seq2-3' (pappu403.key)
1 CNGGTCGACTGATCTAGCTGGGGCCAGACCGGCGACAGGAGTTACCACTACTGA
39 44
61 CAGACAGCGCGATCGAGCCAAACGGTAGTGAGA
...
145 CAACGTGATATATGAAAGTCCGGCTGCTGACTCATTTGTATGCCGCGACCGAGGATTCG
195 200
205 ACAATTAGCTGACTGACTTGAACGAGTCCGAGTGGCTTAACACCAAGCTGCAGAGCT
226 231 238 243
265 GCACGAGAACTGGAAAGTAGTTCA
...
16 matches found in sequence:
aat49166; Partial DNA clone HindPI#2-92 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49166 check: 2309 from: 1 to: 1306
ID AAT49166 standard; DNA: 1306 BP.
XX

```

```

AC  AAT49166:
XX
XX
DE  16-OCT-1997 (first entry)
XX
XX  Partial DNA clone HinfI#2-92 encoding immunostimulatory peptide.
DE
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM  tuberculin skin test; ds.
XX
OS  Mycobacterium tuberculosis.
XX  WO9700067-A1.
XX  PN
XX  PD  03-JAN-1997.
XX  PF  14-JUN-1996; 96WO-US10375.
XX  PR  15-JUN-1995; 95US-0000254.
XX
XX  (UUVI-) UNIV VICTORIA.
PA
XX  Nano FE;
PI
XX  WPI; 1997-077347/07.
DR
XX
XX  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
XX  Claim 1: Page 65-66; 79pp; English.
PS
XX  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells. This
CC  sequence is similar to M. tuberculosis ORF MTCY190.11C. Its encoded
CC  protein is similar to cytochrome C oxidase subunit II.
XX
XX  Sequence 1306 BP; 251 A; 376 C; 407 G; 271 T; 1 other;
SQ
AAT49166 Length: 1306 March 5, 2002 14:19 Type: N Check: 2309 ..
Found using 'seq2-3' (pepu403.key)
1  GTGATACAGAGGCGCCACACGACGACCTCCGCGGCCGACCTGCTTGCACACGCTGTCG
12 17 41 46 50 55
61  CAGTGCAGGCGCTCAGCGCGGCTCCGAGAGGCGCTGCCGTGTCTTCACAGCTGGCGCTC
121  GCAGCAATGCTGGGGGCAATTGGCCGTCACCGTCAGTGAATGACGCTGGT
200  GCATTACCCCGGAGGACACACCTCAATCGAGAAGCTGTGATCGGGGCGGTGACCTCCC
260  TGGCGGTTGGGGTAATGCTGTGGGGTCTCATCTTCTGTCGCGCGGT

```

```

...
643  CGCGGGGCTCAACACCGAGAGACCGGACCTACCTGAATTTGACAGAGTGCAGACGTTGGGC
703  ACCAGCACCGAATTCGGGTGCTGTGTGTCGTCGCGTCCGGACGCAAGCGTA
...
803  TCTTGTCAAGCGTGAAGTGATGCTTAACCCGGGTGGCAACAACACTCGCTCAACGCTTCC
863  AGATCGAAGAAATCACCAGAGACCGGAGCATTCGTGGGCCACTGCGC
...
917  GTGGCAGGTATCACTGATGATGATGACTTCGAGGTCGCGGTCGTGACCCCAAGATTCA
977  AGGCTTACTGTCAGACAGCATCGAGCGGAACAACACCGCCGCTCGGGCGATCA
1037  ACCAGCGCGCCCTTGGCGGTGACCAACCCACCCGTTTGTACTCGCGCGGTGAATTGCGCC
1097  CGCAGCGCGGTAGAGCTTAGACGCTCATGATGAAACCCGACTGTTTGATTTGCTGCGC
1157  GCGTTCTTGTGTGACGCGGCTGTGTACGGGTGTTGACCTGATGTTGCCACCGGT
1217  GGTGTGAGTGGGCTGCGACCACTGCGCTGCGCTTACCGGCGGCGATGGGTTGATGTC
1277  GCCACCTTCTTCGGGTTGTGGCCGCGAT
1279
-----
7 matches found in sequence:
aatt49167: Partial DNA clone AclI#2-823 encoding immunostimulatory peptide.
(from "Mycobactering.seq")
TOIG of: aatt49167 check: 498 from: 1 to: 759
ID  AAT49167 standard; DNA; 759 BP.
XX
XX  AAT49167:
XX
XX  10-NOV-1997 (first entry)
XX
XX  Partial DNA clone AclI#2-823 encoding immunostimulatory peptide.
DE
XX  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KM  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KM  tuberculin skin test; ds.
XX
XX  Mycobacterium tuberculosis.
XX  OS
XX  WO9700067-A1.
XX  PN
XX  PD  03-JAN-1997.
XX  PF  14-JUN-1996; 96WO-US10375.
XX

```



```
207  CWNNTCCCGCGTCCACGCTGCGCTGGTGAGTACGCGCCGACCCCAACGAATGATCGA      |-----|
      238 243
267  CATGGCTGTGGGTAGATGACCGCGC
...
508  AGGCTGGGATGACACCGATGTGCGGCGGCACCGCTCGAGNACGAGTACGCGTGTGCC      |-----|
      558 563
568  CACACNACGGGACCATCGGACGCTTGATCACAACGGGGAGACGGCGCAATGAATCC
      |-----|
628  GCGATCGGCGCTGCGAAATCCGTTGTCTCATNGCACACGGTNAACGAGTTCACCGTGT
      634 639      637 642
688  GCCGC
...
8 matches found in sequence:
aat49169 : Partial DNA clone HinfI#1-3 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49169 check: 8509 from: 1 to: 799

ID  AAT49169 standard; DNA: 799 BP.
XX
AC  AAT49169;
XX
DT  16-OCT-1997 (first entry)
XX
DE  Partial DNA clone HinfI#1-3 encoding immunostimulatory peptide.
XX
KW  Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW  stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW  tuberculin skin test; ds.
XX
OS  Mycobacterium tuberculosis.
XX
PN  M09700067-A1.
XX
PD  03-JAN-1997.
XX
PF  14-JUN-1996; 96MO-US10375.
XX
PR  15-JUN-1995; 95US-0000254.
XX
PA  (UYVI-) UNIV VICTORIA.
XX
PI  Nano FE;
XX
DR  WPI: 1997-077347/07.
XX
PT  New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT  useful in vaccines, diagnostic skin test, immunoassay and gene
PT  isolation
XX
PS  Claim 1: Page 68: 79pp: English.
XX
CC  AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC  which encode partial sequences of immunostimulatory peptides. Each of
CC  the clones encode at least one immunostimulatory T cell epitope. The
CC  clones were identified by testing over 300 fusion clones (alkaline
CC  phosphatase-M. tuberculosis peptide fusions) for their ability to
CC  stimulate interferon (IFN)-gamma production. 80 clones were initially
CC  designated to have some ability to stimulate IFN-gamma production, of
CC  which 76 are shown in AAT49100-175. These sequences can be used to
CC  obtain the full length M. tuberculosis genes and corresponding proteins
```

```
CC  using standard techniques. The peptides are useful in vaccines, as
CC  reagents in an improved tuberculin skin test (especially using peptides
CC  different from those used in vaccines so as to allow differentiation
CC  between vaccinated and infected subjects) and as immunoassay reagents
CC  for detecting specific antibodies. An advantage of these peptides is
CC  that they stimulate production of IFN-gamma (critical for a protective
CC  immune response to M. tuberculosis) by CD4-positive T cells. This
CC  sequence is similar to M. leprae DNA sequence within chromosome region
CC  B983.
XX
SQ  Sequence 799 BP; 149 A; 223 C; 228 G; 129 T; 70 other:
...
AAT49169 Length: 799 March 5, 2002 14:19 Type: N Check: 8509
Found using 'seq2-3' (peppu403.key)

208  CTGCAGCAGTGTCTACGCGGCTTGCCCAAGAACCGGATNCCGANTSGNGGCGCCCATTT      |-----|
      258 263
268  CGCCGCTGNGCTCGACGACGAGATCTTWCGAACGTGCGAATTSGCGCGCGCCG      |-----|
      281 286      284 289
328  TGCAGAGCAKCTTGAAACGCGCGCGCTGCTACCGAGCTGACACAGCAAGGCCN      |-----|
      345 350
388  ANGCTCAASAACG
...
441  CGCGCTGGCAGTTACGAGCANTTCTGTTCAACATCTASTSTGTGCTGCTACGATSAAG      |-----|
      491 496
501  ATCAACGAGCGCGCGCGGACANTCNTGCTGCCGATGCGCGCGGAGGANTCCC      |-----|
      519 524
561  AGCAAGGGGAGGTG
...
616  GCSACCCGTCGCGACCGGNCATCTTCGGCTGTGTCNTGTGATCNCNTGCTGCTCGA      |-----|
      666 671
676  TSGNCATTCGGCTACAGCGGGTTCCTKTCGTGGCCACAKKKAANAACCTACGACGCTAT
      |-----|
736  TTACCGGAGCGCGGNGGAGATCACCCCGGTAACGTGTTATGTTSTCGGCGCTCAAGGTG
      742 747
796  GG
...
8 matches found in sequence:
aat49170 : Partial DNA clone AclI#2-898 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49170 check: 4458 from: 1 to: 713

ID  AAT49170 standard; DNA: 713 BP.
XX
AC  AAT49170;
XX
DT  16-OCT-1997 (first entry)
XX
```

DE Partial DNA clone AcII#2-898 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09700067-A1.
 XX
 PD 03-JAN-1997.
 XX
 PE 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI: 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 68; 79pp; English.
 XX
 CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells. This
 CC sequence is AcII#2-827 translation strand. The AcII#2-827 protein has
 CC similarity to cytochrome d oxidase.
 XX
 SQ Sequence 713 BP; 122 A; 215 C; 178 G; 146 T; 52 other;
 AAT49170 Length: 713 March 5, 2002 14:19 Type: N Check: 4458 ..
 Found using 'seq2-3' (pappu403.key)

1 CTAYCGCAANGCTKNGCAGCGCTGCACNGCAGAAATCGCGGTGCACCCACAGAT
 20 25
 61 TGGCAGTAGCCGGG
 ...
 101 CGTAGCAAAATCANTCGGCCCCGTAATATCTCCGAGATGACAGATGAATGTCCTGCACA
 151 156
 161 TTTCNGNGGTGGCAGTTCCGATACACCGCTSTATCACTTNCAM
 ...
 222 ACCMWCCTGCGCNCNCKTRKSTRKANYRCGAGNTCNAYCAAACTGCTGTGCTCTCACCG
 272 277
 |-----|
 |-----|

282 ATANCCCGCCTGGTATCGCTACACCNAATTTCTGGCAATTTGCTGTGNATCNACN
 297 302
 342 TTTCGCATCGCGCGGCGAGATCGTGCAGNAATRTCAAGTTCCGATGAAGTGGAG
 356 361
 402 CGAGTACTCCGATTCGTGCGGATGCTGCGCGCCCCCGTGGCCATGAGANSCTGGCG
 421 426 433 438
 462 GCCTTCTTCTTGAATCACCCTTCAT
 ...
 503 TCGGCTGAGACAGGTGCGCCCGCGTGGTGCANTCTNGGCTTGATCTGSMATCGTCGGA
 553 558
 563 TNGCGNGTNCACGCTGCCGCTTCTTCATCATCGCNGCAAACT
 ...
 4 matches found in sequence:
 aat49171: Partial DNA clone AcII#2-834 encoding immunostimulatory peptide.
 (from "mycobactereng.seq")
 TOIG of: aat49171 check: 6755 from: 1 to: 274

 ID AAT49171 standard; DNA; 274 BP.
 XX
 AC AAT49171;
 XX
 DT 16-OCT-1997 (first entry)
 XX
 DE Partial DNA clone AcII#2-834 encoding immunostimulatory peptide.
 XX
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
 KW tuberculin skin test; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09700067-A1.
 XX
 PD 03-JAN-1997.
 XX
 PE 14-JUN-1996; 96WO-US10375.
 XX
 PR 15-JUN-1995; 95US-0000254.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI Nano FE;
 XX
 DR WPI: 1997-077347/07.
 XX
 PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
 PT useful in vaccines, diagnostic skin test, immunoassay and gene
 PT isolation
 XX
 PS Claim 1; Page 69; 79pp; English.
 XX
 CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
 CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as

CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.

XX Sequence 274 BP; 50 A; 90 C; 84 G; 50 T; 0 other;

AA149171 Length: 274 March 5, 2002 14:19 Type: N Check: 6755 ..
 Found using 'seq2-3' (pappu403.key)

1 CCGCAGCAGCAGGCAAGCATCGCACCCGTCGATCCCGCATCCCGCGACATGATGTG 47 52

61 CATGTCCGACACCGCCGACCTCGCTTCCGAGTTGACCGCGCTGCGCGTGAGCGC 116

121 CGCAAGCATGCGTCGTTTCATTCGATCCCGCTCGAAATGGCCATGCGC 121

188 ATGCTCGGCGCATGATCTCCACTGCCGCACTGATAAATCGGGTCATGTCGTCGCA 238 243

248 GCGCGACAGGGTAGAGCGCATGACCG

6 matches found in sequence:
 aat49172 ; Partial DNA clone Acil#2-874 encoding immunostimulatory peptide.
 (from "mycobacterieng.seg")
 TOIG of: aat49172 check: 6236 from: 1 to: 252

ID AAT49172 standard; DNA: 252 BP.

XX AC AAT49172;

DT 16-OCT-1997 (first entry)

DE Partial DNA clone Acil#2-874 encoding immunostimulatory peptide.

XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KW stimulate; Interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

XX Mycobacterium tuberculosis.

XX OS

XX PN W09700067-A1.

XX PD 03-JAN-1997.

XX PE 14-JUN-1996; 96WO-US10375.

XX PR 15-JUN-1995; 95US-0000254.

XX PA (UYVI-) UNIV VICTORIA.

XX PI Nano FE;

XX WP1; 1997-077347/07.

XX New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -

XX useful in vaccines, diagnostic skin test, immunoassay and gene

XX isolation

XX Claim 1; Page 70; 79pp; English.

XX AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,

CC which encode partial sequences of immunostimulatory peptides. Each of
 CC the clones encode at least one immunostimulatory T cell epitope. The
 CC clones were identified by testing over 300 fusion clones (alkaline
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to
 CC stimulate interferon (IFN) gamma production. 80 clones were initially
 CC designated to have some ability to stimulate IFN-gamma production, of
 CC which 76 are shown in AAT49100-175. These sequences can be used to
 CC obtain the full length M. tuberculosis genes and corresponding proteins
 CC using standard techniques. The peptides are useful in vaccines, as
 CC reagents in an improved tuberculin skin test (especially using peptides
 CC different from those used in vaccines so as to allow differentiation
 CC between vaccinated and infected subjects) and as immunoassay reagents
 CC for detecting specific antibodies. An advantage of these peptides is
 CC that they stimulate production of IFN-gamma (critical for a protective
 CC immune response to M. tuberculosis) by CD4-positive T cells.

XX Sequence 252 BP; 31 A; 74 C; 88 G; 59 T; 0 other;

AA149172 Length: 252 March 5, 2002 14:19 Type: N Check: 6236 ..
 Found using 'seq2-3' (pappu403.key)

1 GTGATCCTTCAGCATTTGATGTCTCGGTTGATCGTGGCAGCATATAAACCGC 25 30 44 49

61 CTGTTGGGGTGGCGCTCTGCTGGGAGCGCACCTGCTGACCGCTGGCGCTACT 121

121 TCACGGCTTCCTGCTGCTGCTAGCGGTGCGCCGACCGCCCTCTGCGCTGTGATGTG 125 130 152 157

181 GCCGAGTCTGTGGGCACTGCTCAAAACGCTGCTGCCCATTCGATATCGTGTAGCGCT 212 217

241 CCTCGCGCCGCA 241

3 matches found in sequence:
 aat49173 ; Partial DNA clone Acil#2-1018 encoding immunostimulatory peptide.
 (from "mycobacterieng.seg")
 TOIG of: aat49173 check: 843 from: 1 to: 160

ID AAT49173 standard; DNA: 160 BP.

XX AC AAT49173;

DT 16-OCT-1997 (first entry)

DE Partial DNA clone Acil#2-1018 encoding immunostimulatory peptide.

XX Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;

KW stimulate; Interferon-gamma; IFN-gamma; production; vaccine;

KW tuberculin skin test; ds.

XX Mycobacterium tuberculosis.

XX OS

XX PN W09700067-A1.

XX PD 03-JAN-1997.

XX PE 14-JUN-1996; 96WO-US10375.

XX PR 15-JUN-1995; 95US-0000254.

XX PA (UYVI-) UNIV VICTORIA.

XX PI Nano FE;

XX WP1; 1997-077347/07.

PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
PS Claim 1: Page 70; 79pp; English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ Sequence 160 BP; 14 A; 57 C; 68 G; 21 T; 0 other;
AAT49173 Length: 160 March 5, 2002 14:19 Type: N Check: 843 ..
Found using 'seq2-3' (pappu403.key)
1 |-----|
1 GGGCGCGCGTGTGCTGTGCGCCGCCGGCGGTGGGGTCCGGCCAGCGTGTCCGCC
6 |-----|
61 AGTGGCCGCCGCGCAAGTATTGGCCGCGCTCCCTCGAGCAGCAGCAGCGGCTCGGGGCG
86 91 104 109
121 GGGCGTGTGCGCGCGCTGCGCAAGCTGCCACCCGGTGG

8 matches found in sequence:
aat49174 : Partial DNA clone Hinf#1-27 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49174 check: 4063 from: 1 to: 393
ID AAT49174 standard; DNA: 393 BP.
XX
AC AAT49174;
XX
DT 16-OCT-1997 (first entry)
XX
DE Partial DNA clone Hinf#1-27 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX tuberculin skin test; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN W09700067-A1.
XX
PD 03-JAN-1997.
XX
PE 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI Nano FE;
XX
KW WPI; 1997-077347/07.
DR

XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX
PS Claim 1: Page 1; 79pp; English.
XX
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,
CC which encode partial sequences of immunostimulatory peptides. Each of
CC the clones encode at least one immunostimulatory T cell epitope. The
CC clones were identified by testing over 300 fusion clones (alkaline
CC phosphatase-M. tuberculosis peptide fusions) for their ability to
CC stimulate interferon (IFN)-gamma production. 80 clones were initially
CC designated to have some ability to stimulate IFN-gamma production, of
CC which 76 are shown in AAT49100-175. These sequences can be used to
CC obtain the full length M. tuberculosis genes and corresponding proteins
CC using standard techniques. The peptides are useful in vaccines, as
CC reagents in an improved tuberculin skin test (especially using peptides
CC different from those used in vaccines so as to allow differentiation
CC between vaccinated and infected subjects) and as immunoassay reagents
CC for detecting specific antibodies. An advantage of these peptides is
CC that they stimulate production of IFN-gamma (critical for a protective
CC immune response to M. tuberculosis) by CD4-positive T cells.
XX
SQ Sequence 393 BP; 72 A; 120 C; 119 G; 82 T; 0 other;
AAT49174 Length: 393 March 5, 2002 14:19 Type: N Check: 4063 ..
Found using 'seq2-3' (pappu403.key)

1 |-----| |-----| |-----|
1 ATCAGCCGGGGTGCAGCCGCCGATGACCTTCGCTGCTGCTGCGTCCGATATC
15 20 33 38 42 47 36 41 45 50
61 AATCCATACCATCTCTTACGCACCTTTAGAGTGTG

130 TTGGGGCTCGCGTTGCGTAGATCTTCCGCCAGCCTCGATGCTCTATCAGACATCA
180 185
190 GTCCGTCGGGCGCAGCAACCCCGCACCGTGGAGTGTGTCAGGT

318 GTAGACGGGGCTGTTCGAGAGGTGTAGTCTAGTCCACAGAGGTTCGAGCCACAG
368 373
378 TTAATACACAGCGTGC

3 matches found in sequence:
aat49175 : Partial DNA clone #2-147 encoding immunostimulatory peptide.
(from "mycobacterieng.seq")
TOIG of: aat49175 check: 9365 from: 1 to: 248
ID AAT49175 standard; DNA: 248 BP.
XX
AC AAT49175;
XX
DT 16-OCT-1997 (first entry)
XX
DE Partial DNA clone #2-147 encoding immunostimulatory peptide.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
XX tuberculin skin test; ds.
XX


```

260 TGGCGTTGGGGTAATCGTGTGGGGTCTCATCTTCTGTGTCGGGGT
...
643 CGCGGGCTCAACACCGAGGACCGGACCTACTGAATTTCAGACAGGTGGGCG
|-----|
703 ACCAGCACCAGAAATTCCGGTGTGCTGCTGCCGTCGGGCAACGGTA
...
803 TCTTTTCAGCGGTGACGTGATGCTTAACCCGGTGGCAACAACCTCGTCAACGCTTCC
|-----|
863 AGATCGAAGAAATACACCAAGACCGGAGCATGCTGGGCACTGGCG 853 858
...
917 GTGCGACGTATCCTCATGATGAATCTTGAGTCCGGTGTGACCCCAAGATTCA
|-----|
977 AGGCCTACTGCGACGACGATCGACGGGATACACGCGCCCTGCGGCGATCA
|-----|
1037 ACCAGCCCGCCCTTGGGGTGACACCAACCGCTTGATATCTGCGCGGTGAATTGGCCC
1012 1017 1029 1034
1097 CGCAGCCCGGTAGTTAGACGCTC
|-----|
1114 1119
-----
1 match found in sequence:
aav21737 : Vlnneo sequence used for construction of synthetic HIV env DNA.
(from "mycobacterieng.seq")
TOIG of: aav21737 check: 8348 from: 1 to: 28
ID AAT49180 standard; DNA: 28 BP.
AC AAT49180;
XX
DT 16-OCT-1997 (first entry)
XX
DE Primer for MTB2-92 ORF amplification.
XX
KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;
KW tuberculin skin test; primer; PCR; ss.
XX
OS Synthetic.
XX
PN WO9700067-A1.
XX
PD 03-JAN-1997.
XX
PE 14-JUN-1996; 96WO-US10375.
XX
PR 15-JUN-1995; 95US-0000254.
XX
PA (UYVT-) UNIV VICTORIA.
XX
PI Nano FE;
XX
DR WPI; 1997-077347/07.
XX
PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -
PT useful in vaccines, diagnostic skin test, immunoassay and gene
PT isolation
XX

```

```

PS Example 2; Page 19; 79pp; English.
XX AAT49179-80 are primers used to amplify a the mb2-92 gene with a
CC histidine tag. The mb2-92 gene was identified based on
CC clone Hinp2-92, which encodes a partial immunostimulatory peptide.
CC AAT49100-175 comprise M. tuberculosis cDNA clones which encode at least
CC one immunostimulatory T cell epitope. The clones were identified by
CC testing over 300 fusion clones (alkaline phosphatase-M. tuberculosis
CC peptide fusions) for their ability to stimulate interferon (IFN)-gamma
CC production. 80 clones were initially designated to have some ability to
CC stimulate IFN-gamma production, of which 76 are shown in AAT49100-175.
CC These sequences can be used to obtain the full length M. tuberculosis
CC genes and corresponding proteins using standard techniques. The peptides
CC are useful in vaccines, as reagents in an improved tuberculin skin test
CC (especially using peptides different from those used in vaccines so as
CC to allow differentiation between vaccinated and infected subjects) and
CC as immunoassay reagents for detecting specific antibodies. An advantage
CC of these peptides is that they stimulate production of IFN-gamma
CC (critical for a protective immune response to M. tuberculosis) by
CC CD4-positive T cells.
XX
SQ Sequence 28 BP; 5 A; 11 C; 8 G; 4 T; 0 other;
AAT49180 Length: 28 March 5, 2002 14:19 Type: N Check: 8348 ..
Found using 'seq2-3' (pappu403.key)
1 AAGCTTCGCCATGCCCGGTAAAGCGCC
|-----|
23 28
-----
29 matches found in sequence:
aav21737 : Vlnneo sequence used for construction of synthetic HIV env DNA.
(from "mycobacterieng.seq")
TOIG of: aav21737 check: 5515 from: 1 to: 4864
ID AAV21737 standard; DNA: 4864 BP.
XX
AC AAV21737;
XX
DT 20-JUL-1998 (first entry)
XX
DE Vlnneo sequence used for construction of synthetic HIV env DNA.
XX
KW Vlnneo; HIV vaccine; codon usage optimisation; HIV; immune response;
KW human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KW env protein; recombinant; ds.
XX
OS Synthetic.
XX
PN WO9748370-A2.
XX
PD 24-DEC-1997.
XX
PE 17-JUN-1997; 97WO-US10517.
XX
PR 16-JUL-1996; 96GB-0014943.
XX
PR 21-JUN-1996; 96US-0020165.
XX
PR 21-JUN-1996; 96US-0020166.
XX
PR 16-JUL-1996; 96GB-0014942.
XX
PA (MERI) MERCK & CO INC.
XX
PI Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW;
XX
DR WPI; 1998-062825/06.
XX
PT DNA encoding protein with codon usage optimised for intended host
PT cell - specifically for DNA vaccines against human immune deficiency
PT virus, allows rev-independent expression of HIV genes
XX
PS Example 4; Pages 73-76; 113pp; English.
XX

```

This VJneo sequence is used in the construction of a new synthetic DNA, encoding a protein or peptide. The synthetic DNA sequence encodes HIV env protein or a fragment and has codons optimised for expression in a non-homologous host. The synthetic DNA sequences are used to increase production of recombinant protein and to induce anti-HIV neutralising antibody, HIV-specific T-cell immune responses or protective immune responses in vertebrates, specifically as HIV vaccines. Optimisation of codons results in increased expression of the DNA in the host. The DNA induces antigen-presenting cells to stimulate cytotoxic and helper T-cells, and effector functions such as lymphokine secretion specific to HIV antigens. Cross-strain protection is achieved without use of adjuvant and the synthetic DNA may provide a longer-lasting response. The env gene includes a high proportion of codons rarely used in mammals, so optimisation will allow this gene to be expressed efficiently in absence of rev. The DNA sequences, more generally, may express antigens from many other pathogens, e.g. hepatitis or tuberculosis, and also from tumours.

Sequence 4864 BP; 1229 A; 1226 C; 1168 G; 1241 T; 0 other;

AAV21737 Length: 4864 March 5, 2002 14:19 Type: N Check: 5515 ..
Found using 'seq2-3' (pappu403.key)

```
390 AGTCCCGCTTACATACTAGCGTAATGGCGCCCTGGCTGACCGCCCAAGCAGCCGCC
      |-----|
450 GCCCATGACGTCATATATGAGCATATGTTCCCATAGTACGCCCAATAGGAGCTTTCATTT
      457 462 |-----| 487 492
510 GACGTCATGGGTGAGTATTACGTAACCTGCCACTTGGCAGTACATCAAGTATTC
      510 515 |-----|
570 ATATGCCAAGTAGCGCCCTATTTGACGTCAATGAGCGTAATGGCCCGCTGGCATATAG
      593 598 |-----|
630 CCCAGTACATGACCTTATGAGGACTTCTCTACTTGGCAGTACATCTACGTAATGATCAGC
      686 |-----|
690 CTATTACCATGTGATCGCGTTTGGCAGTACATCAATGCGCGTGAGTAGCGGTTTGACT
      691
750 CACGGGATTTCCAAAGTCTCCACCCCATTTGACGTCAATGGGAGTTGTTTTGGCACAA
      779 784 |-----|
810 ATCAACGGGACTTTCAAAATGTCG
      923 928 |-----|
873 GTGTACGGTGGAGGCTTATATTAAGCAGAGCTCGTTAGTAACCGTCAGATCGCTGGA
      933 938 |-----|
933 GACGGCATCCACGCTGTTTGGACCTCCATAGAAAGACACGGGACCGATCCAGCCTC
      933 938 |-----|
1159 TAGCCTATAGGTGGGTATTATGAACATTATTGACCACTCCCGTATTGGTACGATTA
      1209 1214 |-----|
1219 TCCATTACTAATCCATAACATGGCTCTTTGCCACAACCTCTCTTAT
      1209 1214 |-----|
1452 TCTCGGTAGCGGGGAGACTTCTACATCCGAGCCCTGCTCCATGCTTCAGCCACTAT
      1502 1507 |-----|
1512 GGTGCTCGGAGAGCTCTTGGCTCTACAGTGGAGGCCAGACTTAGGCACAGCAGCA
      1513 1518 |-----|
2417 GAATCATAGAAATTCTTCCGCTTCCTGCTCAGTACGCTGCGTGGCGGCTGCGGC
      2467 2472 |-----|
2477 TCGGCGAGCGGTATCAGCTACCTCAAGAGCGGTATACGGTTATC
      2607 2612 |-----|
2557 TGAGCAAAAGGCCAGCAAAAGGCCAGGAACGTTAAAGGCCGCTTCTGCGCTTTTC
      2654 2659 |-----|
2617 CATAGGCTCCGCCCCCTGACGAGCATCACAAAAATGACGCTCAAGTCAGAGTGGGCA
      2703 2708 |-----|
2677 AACCCGACAGACATATAAAGATACGAGCGCTTCCCGCTGGAAGCTCCCTGCGCTCT
      2796 |-----|
2737 CCTGTTCCGACCCCTGCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCCGGAAGCGT
      2801 2804 |-----|
2797 GCGCTTCTCAATGCTCAACGCTGATGATATCTAGTTGCGGTAGCTGCTGCTCCAG
      2842 2847 |-----|
2857 CTGGGCTGTGACAGAAACCCCGCTCAGCCGACCGCTGCGCTTATCCGGTAACAT
      2915 |-----|
2917 GGTCTTGAGTCCACCCGCTAAGACAGACTTATGCGCAGTGGCAGCAGCACTGGTAAC
      2949 2954 |-----|
2977 AGGATTACAGAGCGAGGTATGTAGCG
      3232 3237 |-----|
3182 GCAGAAAAAAGATCTCAGAGATCTTTGATCTTTTCTAAGGGGTCTGACGCTCAGT
      3405 3410 |-----|
3242 GGAACGAAACTCAGCTTAAGGATTTTGGTCATGAGATATACAA
      3456 3461 |-----|
3355 TAAACTTGTGTGACAGTTTACCAATGCTTAATCACTGAGGACCTATCTCAGGATCTGT
      3509 3514 |-----|
3415 CTATTCTGTATCCATAGTTGCTGACTCCGGGGGGGGGGGGGCTGAGGTCTGCTCG
      3559 3564 |-----|
3475 TGAAGAGGTGTGCTGACTATACAGAGCCTGAATTCGCCCATATCCAGCAGAAAGT
      3599 3604 |-----|
3535 GAGGAGCCACGGTTGATGAGAGCTTTGTT
      3609 3614 |-----|
```

```

...
3928 AATTCCCGCTGTCAAAAATAGTTATCATAGTAGAATAATCACCATGATGACGACTGAA
      |-----|
      3978 3983
3988 TCCGGTGAAGATGCCAAAAGCTTATGCAATTCTTCCAGACTTGT
...
4078 ACCAAACCGTTATTCATTCGTGATTCGCCCTGAGCGAGACGAATAACCGGATCGCTGTA
      |-----|
      4128 4133
4138 AAGGACAAATTACAACAGAAATCGAATGCAACCGCGCAGGAGACA
...
4327 AATTCCGTCAGCCAGTTAGTCTGACCATCTCATCTGTAACATCATTTGGCAACGCTACT
      |-----|
      4377 4382
4387 TTGCCATGTTTCAGAAACACTCTGGCGCATCGGCGCTTCCATACA
...
4484 ACCCATATAAATCAGCATCATCTTGAATTTAATCGCGGCGCTCGAGCAAGCTTCC
      |-----|
      4534 4539
4544 GTTGATATGGCTCATTAACACCCCTTGTTACTGTTTATGTAGAC
...
4745 AAAATAAACAAATAGGGGTCCGGGCACATTTCCCGAAAGTGGCACCCTGCTAG
      |-----|
      4795 4800
4805 AAACCATTTATTCATCATGACATTAACTATMAAATATAGCGCTATCAC
...
-----
1 match found in sequence:
aav21741 ; HIVm gp120 gene amplifying antisense primer.
(from "mycobacterng.seq")
TOIG of: aav21741 check: 6495 from: 1 to: 48
ID AAV21741 standard; DNA: 48 BP.
XX
XX AAV21741;
AC
XX
XX 20-JUL-1998 (first entry)
DT
XX
XX HIVm gp120 gene amplifying antisense primer.
DE
XX
XX Vlnco; HIV vaccine; codon usage optimisation; HIV; immune response;
KW human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KW env protein; gp120 gene; recombinant; PCR primer; ss.
XX
OS Synthetic.
OS Human immunodeficiency virus.
OS
XX
XX PN W09748370-A2.
XX
XX PD 24-DEC-1997.
XX
XX PF 17-JUN-1997; 97WO-US10517.
XX
XX PR 16-JUL-1996; 96GB-0014943.
XX PR 21-JUN-1996; 96US-0020165.
XX PR 21-JUN-1996; 96US-0020166.
XX PR 16-JUL-1996; 96GB-0014942.

```

```

XX
XX (MERI ) MERCK & CO INC.
XX
XX PI Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW;
XX
XX DR WPI: 1998-062825/06.
XX
XX PT DNA encoding protein with codon usage optimised for intended host
XX PT cell - specifically for DNA vaccines against human immune deficiency
XX PT virus, allows rev-independent expression of HIV genes
XX
XX PS Example 6; Page 38; 113pp; English.
XX
XX CC This primer is used for the PCR amplification of HIVm gp120 gene
XX CC deleting the first 30 amino acids of the peptide leader sequence. This
XX CC is used to construct a Vlnco expression vector for vaccine production
XX CC containing a new synthetic DNA, encoding a protein or peptide. The
XX CC synthetic DNA sequence encodes HIV env protein or its fragment and has
XX CC codons optimised for expression in a non-homologous host. The synthetic
XX CC DNA sequences are used to increase production of recombinant protein and
XX CC to induce anti-HIV neutralising antibody, HIV-specific T-cell immune
XX CC responses or protective immune responses in vertebrates, specifically as
XX CC HIV vaccines. Optimisation of codons results in increased expression of
XX CC the DNA in the host. The DNA induces antigen-presenting cells to
XX CC stimulate cytotoxic and helper T-cells, and effector functions such as
XX CC lymphokine secretion specific to HIV antigens. Cross-strain protection is
XX CC achieved without use of adjuvant and the synthetic DNA may provide a
XX CC longer-lasting response. The env gene includes a high proportion of
XX CC codons rarely used in mammals, so optimisation will allow this gene to be
XX CC expressed efficiently in absence of rev. The DNA sequences, more
XX CC generally, may express antigens from many other pathogens, e.g. hepatitis
XX CC or tuberculosis, and also from tumours.
XX
XX SQ Sequence 48 BP; 7 A; 20 C; 6 G; 15 T; 0 other;
XX
XX AAV21741 Length: 48 March 5, 2002 14:19 Type: N Check: 6495 ..
Found using 'seq2-3' (pappu403.key)
1
1 CCCCAGGAATCCACCTGTTAGCGCTTTCCTCTCTGACACACTCTCTC
20 25
-----
1 match found in sequence:
aav21747 ; HIV gp30-opt DNA amplifying antisense primer.
(from "mycobacterng.seq")
TOIG of: aav21747 check: 8453 from: 1 to: 54
ID AAV21747 standard; DNA: 54 BP.
XX
XX AC AAV21747;
XX
XX DT 20-JUL-1998 (first entry)
XX
XX DE HIV gp30-opt DNA amplifying antisense primer.
XX
XX KW Vlnco; HIV vaccine; codon usage optimisation; HIV; immune response;
KW human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KW env protein; recombinant; PCR primer; ss.
XX
XX OS Synthetic.
XX OS Human immunodeficiency virus.
XX OS
XX
XX PN W09748370-A2.
XX
XX PD 24-DEC-1997.
XX
XX PF 17-JUN-1997; 97WO-US10517.
XX
XX PR 16-JUL-1996; 96GB-0014943.
XX PR 21-JUN-1996; 96US-0020165.
XX PR 21-JUN-1996; 96US-0020166.
XX PR 16-JUL-1996; 96GB-0014942.

```

```
XX (MERI ) MERCK & CO INC.
PA Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW.
XX WPI; 1998-062825/06.
XX
XX DNA encoding protein with codon usage optimised for intended host
PT cell - specifically for DNA vaccines against human immune deficiency
PT virus, allows rev-independent expression of HIV genes
XX
XX Example 7; Page 41; 113pp; English.
PS
XX This primer is used for the PCR amplification of the HIV gp30-opt DNA
CC from gp32-opt DNA. This is used in the construction of a VJneo
CC expression vector for vaccine production containing a new synthetic DNA,
CC encoding a protein or peptide. The synthetic DNA sequence encodes HIV
CC env protein or its fragment and has codons optimised for expression in a
CC non-homologous host. The synthetic DNA sequences are used to increase
CC production of recombinant protein and to induce anti-HIV neutralising
CC antibody. HIV-specific T-cell immune responses or protective immune
CC responses in vertebrates, specifically as HIV vaccines. Optimisation
CC of codons results in increased expression of the DNA in the host. The
CC DNA induces antigen-presenting cells to stimulate cytotoxic and helper
CC T-cells, and effector functions such as lymphokine secretion specific
CC to HIV antigens. Cross-strain protection is achieved without use of
CC adjuvant and the synthetic DNA may provide a longer-lasting response. The
CC env gene includes a high proportion of codons rarely used in mammals, so
CC optimisation will allow this gene to be expressed efficiently in absence
CC of rev. The DNA sequences, more generally, may express antigens from many
CC other pathogens, e.g. hepatitis or tuberculosis, and also from tumours.
XX
SQ Sequence 54 BP; 12 A; 13 C; 13 G; 16 T; 0 other;
AAV21747 Length: 54 March 5, 2002 14:19 Type: N Check: 8453 ..
Found using 'seq2-3' (pappu403.key)
1
1 CCACATGATATGCGCGGCTTATTGATGATGACACAGCCAGTGTGATG
10 15
1 match found in sequence:
aav21760; HIV gp143 intracellular peptide domain generating antisense primer.
(from "mycobacterieng.seq")
TOIG of: aav21760 Check: 8739 from: 1 to: 62
ID AAV21760 standard; DNA: 62 BP.
XX
XX AAV21760;
AC
XX 20-JUL-1998 (first entry)
DT
XX
XX HIV gp143 intracellular peptide domain generating antisense primer.
DE
XX
XX VJneo; HIV vaccine: codon usage optimisation; HIV; immune response;
KM human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KW env protein; gp143; recombinant; PCR primer; ss.
XX
XX Synthetic.
OS Human immunodeficiency virus.
XX
XX WO9748370-A2.
PN
XX
XX 24-DEC-1997.
PD
XX
XX 17-JUN-1997; 97WO-US10517.
PF
XX
XX 16-JUL-1996; 96GB-0014943.
PR 21-JUN-1996; 96US-0020165.
PR 21-JUN-1996; 96US-0020166.
PR 16-JUL-1996; 96GB-0014942.
XX
```

```
PA (MERI ) MERCK & CO INC.
XX Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW.
XX WPI; 1998-062825/06.
XX
XX DNA encoding protein with codon usage optimised for intended host
PT cell - specifically for DNA vaccines against human immune deficiency
PT virus, allows rev-independent expression of HIV genes
XX
XX Example 10; Page 49; 113pp; English.
PS
XX This PCR primer is used for replacing the intracellular peptide domain
CC of gp143 with that of glycoprotein B. This is used to construct a VJneo
CC expression vector for vaccine production containing a new synthetic DNA,
CC encoding a protein or peptide. The synthetic DNA sequence encodes HIV
CC env protein or its fragment and has codons optimised for expression in a
CC non-homologous host. The synthetic DNA sequences are used to increase
CC production of recombinant protein and to induce anti-HIV neutralising
CC antibody. HIV-specific T-cell immune responses or protective immune
CC responses in vertebrates, specifically as HIV vaccines. Optimisation
CC of codons results in increased expression of the DNA in the host. The
CC DNA induces antigen-presenting cells to stimulate cytotoxic and helper
CC T-cells, and effector functions such as lymphokine secretion specific
CC to HIV antigens. Cross-strain protection is achieved without use of
CC adjuvant and the synthetic DNA may provide a longer-lasting response. The
CC env gene includes a high proportion of codons rarely used in mammals, so
CC optimisation will allow this gene to be expressed efficiently in absence
CC of rev. The DNA sequences, more generally, may express antigens from many
CC other pathogens, e.g. hepatitis or tuberculosis, and also from tumours.
XX
SQ Sequence 62 BP; 15 A; 19 C; 15 G; 13 T; 0 other;
AAV21760 Length: 62 March 5, 2002 14:19 Type: N Check: 8739 ..
Found using 'seq2-3' (pappu403.key)
1
1 CCACATGATATGCGCGGCTTATTAGCGCTTGTATGACGCGGTTACACATGACACAGACA
10 15
1 match found in sequence:
aav21763; HIV MN VI-V5 (opt) gene segment.
(from "mycobacterieng.seq")
TOIG of: aav21763 Check: 751 from: 1 to: 1065
ID AAV21763 standard; DNA: 1065 BP.
XX
XX AAV21763;
AC
XX 20-JUL-1998 (first entry)
DT
XX
XX HIV MN VI-V5 (opt) gene segment.
DE
XX
XX VJneo; HIV vaccine: codon usage optimisation; HIV; immune response;
KM human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KW env protein; recombinant; ss.
XX
XX Synthetic.
OS Human immunodeficiency virus.
XX
XX WO9748370-A2.
PN
XX
XX 24-DEC-1997.
PD
XX
XX 17-JUN-1997; 97WO-US10517.
PF
XX
XX 16-JUL-1996; 96GB-0014943.
PR 21-JUN-1996; 96US-0020165.
PR 21-JUN-1996; 96US-0020166.
PR 16-JUL-1996; 96GB-0014942.
XX
```



```
XX VLJneo; HIV vaccine; codon usage optimisation; HIV; immune response;
KW human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KM env protein; rev response element; ss.
OS Synthetic.
OS Human immunodeficiency virus type 1.
PN WO9748370-A2.
XX
XX 24-DEC-1997.
XX
XX 17-JUN-1997; 97WO-US10517.
XX
XX 16-JUL-1996; 96GB-0014943.
XX 21-JUN-1996; 96US-0020165.
XX 21-JUN-1996; 96US-0020166.
XX 16-JUL-1996; 96GB-0014942.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Davies ME, Freed DC, Liu MA, Perry HC, Shilver JW;
XX WPI; 1998-062825/06.
XX
XX DNA encoding protein with codon usage optimised for intended host
XX cell - specifically for DNA vaccines against human immune deficiency
XX virus, allows rev-independent expression of HIV genes
XX
XX Example 11; Page 54; 113pp; English.
XX
XX This DNA segment RRE.Mut (B) corresponds to HIV-1 rev response element
XX and has optimal codon usage for expression. This can be used in an
XX expression vector for vaccine production. The expression vector contains
XX a new synthetic DNA, encoding a HIV env protein or its fragment and has
XX codons optimised for expression in a non-homologous host. The synthetic
XX DNA sequences are used to increase production of recombinant protein and
XX to induce anti-HIV neutralising antibody, HIV-specific T-cell immune
XX responses or protective immune responses in vertebrates, specifically as
XX HIV vaccines. Optimisation of codons results in increased expression of
XX the DNA in the host. The DNA induces antigen-presenting cells to
XX stimulate cytotoxic and helper T-cells, and effector functions such as
XX lymphokine secretion specific to HIV antigens. Cross-strain protection is
XX achieved without use of adjuvant and the synthetic DNA may provide a
XX longer-lasting response. The env gene includes a high proportion of
XX codons rarely used in mammals, so optimisation will allow this gene to
XX be expressed efficiently in absence of rev. The DNA sequences, more
XX generally, may express antigens from many other pathogens, e.g. hepatitis
XX or tuberculosis, and also from tumours.
XX
XX Sequence 354 BP; 82 A; 96 C; 113 G; 63 T; 0 other:
XX
AAV21765 Length: 354 March 5, 2002 14:19 Type: N Check: 8930 ..
Found using 'seq2-3' (pappu403.key)
..
113 GCATTGGGGCCCTGTTTCTTG6GCTTTCGTGGGGGCTGCTGCCCAACAAATGGGCGCCCTGA
|-----|
163 168
173 GCATGACCCCTACCCGTCGAAGCTGCGCAGCTGCTGAGTGGCATGGCATGACGACGAGCAACAA
|-----|
214 219
233 ACCTGCTCCGGCGGCATCGAAGCCGACGACGACCTCCT
...
273 GCTGACTGTGTGGGGGATCAAAACAGCTTCAGAGCCCGGGTGCTGCGCGCTGATCT
|-----|
323 328
```

```
333 GAAAGACCGCACTCTAGGC
-----
1 match found in sequence:
aav21766 : HIV gp32 (opt) gene segment from AVR11 site.
(from "mycobacterieng.seq")
TOIG of: aav21766 check: 2953 from: 1 to: 387
ID AAV21766 standard; DNA; 387 BP.
XX
XX AAV21766;
XX
XX 20-JUL-1998 (first entry)
XX
XX HIV gp32 (opt) gene segment from AVR11 site.
XX
XX VLJneo; HIV vaccine; codon usage optimisation; HIV; immune response;
KW human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KM env protein; gp32; recombinant; ss.
XX
XX Synthetic.
XX OS Human Immunodeficiency virus.
XX
XX WO9748370-A2.
XX
XX 24-DEC-1997.
XX
XX 17-JUN-1997; 97WO-US10517.
XX
XX 16-JUL-1996; 96GB-0014943.
XX 21-JUN-1996; 96US-0020165.
XX 21-JUN-1996; 96US-0020166.
XX 16-JUL-1996; 96GB-0014942.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Davies ME, Freed DC, Liu MA, Perry HC, Shilver JW;
XX WPI; 1998-062825/06.
XX
XX DNA encoding protein with codon usage optimised for intended host
XX cell - specifically for DNA vaccines against human immune deficiency
XX virus, allows rev-independent expression of HIV genes
XX
XX Example 11; Pages 54-55; 113pp; English.
XX
XX This is a HIV gp32 gene segment from the AVR11 site to the end of gp143
XX having optimal codon usage (opt) for expression. This can be used in an
XX expression vector for vaccine production. The expression vector contains
XX a new synthetic DNA, encoding a HIV env protein or its fragment and has
XX codons optimised for expression in a non-homologous host. The synthetic
XX DNA sequences are used to increase production of recombinant protein and
XX to induce anti-HIV neutralising antibody, HIV-specific T-cell immune
XX responses or protective immune responses in vertebrates, specifically as
XX HIV vaccines. Optimisation of codons results in increased expression of
XX the DNA in the host. The DNA induces antigen-presenting cells to
XX stimulate cytotoxic and helper T-cells, and effector functions such as
XX lymphokine secretion specific to HIV antigens. Cross-strain protection is
XX achieved without use of adjuvant and the synthetic DNA may provide a
XX longer-lasting response. The env gene includes a high proportion of
XX codons rarely used in mammals, so optimisation will allow this gene to
XX be expressed efficiently in absence of rev. The DNA sequences, more
XX generally, may express antigens from many other pathogens, e.g. hepatitis
XX or tuberculosis, and also from tumours.
XX
XX Sequence 387 BP; 91 A; 102 C; 113 G; 81 T; 0 other:
XX
AAV21766 Length: 387 March 5, 2002 14:19 Type: N Check: 2953 ..
Found using 'seq2-3' (pappu403.key)
...
|-----|
```

329 CTGTCATTGTGACCGGGTGAACAGAGGCTACTCCCTAATTAAGCCGGCGATATC
379 384

3 matches found in sequence:

aav21767 : Simian retrovirus-1 3'-UTR synthetic gene segment SRV-1 CTE (A).
(from "mycobactereng.seq")
TOIG of: aav21767 check: 1027 from: 1 to: 269

ID AAV21767 standard; DNA; 269 BP.

XX AAV21767;

XX 20-JUL-1998 (first entry)

XX Simian retrovirus-1 3'-UTR synthetic gene segment SRV-1 CTE (A).

KW Vlnreo; HIV vaccine; codon usage optimisation; HIV; immune response;
human immunodeficiency virus; hepatitis; tuberculosis; tumour;

KM env protein; gp120; recombinant; ss.

XX Synthetic.

OS Simian retrovirus sp.

XX WO9748370-A2.

XX 24-DEC-1997.

XX 17-JUN-1997; 97WO-US10517.

XX 16-JUL-1996; 96GB-0014943.

PR 21-JUN-1996; 96US-0020165.

PR 21-JUN-1996; 96US-0020166.

PR 16-JUL-1996; 96GB-0014942.

XX (MERI) MERCK & CO INC.

PI Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW;

XX WPI: 1998-062825/06.

PT DNA encoding protein with codon usage optimised for intended host
cell - specifically for DNA vaccines against human immune deficiency
virus, allows rev-independent expression of HIV genes

XX Example 11; Page 55; 113pp; English.

This SRV-1 CTE (A) is a synthetic gene segment corresponding to a 3'-UTR from the simian retrovirus-1 genome. This is placed in the 3' terminus of the HIV genes to increase rev-independent expression. This can be used in an expression vector for vaccine production. The vector contains a new synthetic DNA, encoding a HIV env protein or its fragment and has codons optimised for expression in a non-homologous host. The synthetic DNA sequences are used to increase production of recombinant protein and to induce anti-HIV neutralising antibody, HIV-specific T-cell immune responses or protective immune responses in vertebrates, specifically as HIV vaccines. Optimisation of codons results in increased expression of the DNA in the host. The DNA induces antigen-presenting cells to stimulate cytotoxic and helper T-cells, and effector functions such as lymphokine secretion specific to HIV antigens. Cross-strain protection is achieved without use of adjuvant and the synthetic DNA may provide a longer-lasting response. The env gene includes a high proportion of codons rarely used in mammals, so optimisation will allow this gene to be expressed efficiently in absence of rev. The DNA sequences, more generally, may express antigens from many other pathogens, e.g. hepatitis or tuberculosis, and also from tumours.

XX Sequence 269 BP; 71 A; 66 C; 72 G; 60 T; 0 other;

AAV21767 Length: 269 March 5, 2002 14:19 Type: N Check: 1027 ..
Found using 'seq2-3' (pappu403.key)

I-----I

1 GCCCGGCGATATCTAGACCCTCCCTCGAGCTAAGCTGAGACCAATGACGGTA
6 11

61 A

...

138 ATAAAAATGTACTACCTCAACCTAAGACAGCGCACCTTCGAGGATTTGTCGCTCTT
188 193

198 TTATATATATTTAAAAAGGCTGACTGTCGGAGCCCTGCTGCCCGATGATGCTTTGGGA

258 TATCGCCCGCGC

259 264

3 matches found in sequence:

aav21768 : Simian retrovirus-1 3'-UTR synthetic gene segment SRV-1 CTE (B).
(from "mycobactereng.seq")
TOIG of: aav21768 check: 305 from: 1 to: 269

ID AAV21768 standard; DNA; 269 BP.

XX AAV21768;

XX 20-JUL-1998 (first entry)

XX Simian retrovirus-1 3'-UTR synthetic gene segment SRV-1 CTE (B).

KW Vlnreo; HIV vaccine; codon usage optimisation; HIV; immune response;
human immunodeficiency virus; hepatitis; tuberculosis; tumour;

KM env protein; gp120; recombinant; ss.

XX Synthetic.

OS Simian retrovirus sp.

XX WO9748370-A2.

XX 24-DEC-1997.

XX 17-JUN-1997; 97WO-US10517.

XX 16-JUL-1996; 96GB-0014943.

PR 21-JUN-1996; 96US-0020165.

PR 21-JUN-1996; 96US-0020166.

PR 16-JUL-1996; 96GB-0014942.

XX (MERI) MERCK & CO INC.

PI Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW;

XX WPI: 1998-062825/06.

PT DNA encoding protein with codon usage optimised for intended host
cell - specifically for DNA vaccines against human immune deficiency
virus, allows rev-independent expression of HIV genes

XX Example 11; Page 56; 113pp; English.

This SRV-1 CTE (B) is a synthetic gene segment from the simian retrovirus-1 genome having a single nucleotide mutation associated with increased mRNA turnover compared to SRV-1 CTE (A) (AAV21767). This can be used in an expression vector for vaccine production. The expression vector contains a synthetic DNA, encoding a HIV env protein or its fragment and has codons optimised for expression in a non-homologous host. The synthetic DNA sequences are used to increase production of recombinant protein and to induce anti-HIV neutralising antibody, HIV-specific T-cell immune responses or protective immune responses in vertebrates, specifically as HIV vaccines. Optimisation of codons results in increased expression of the DNA in the host. The DNA induces antigen-presenting cells to stimulate cytotoxic and helper T-cells, and

CC effector functions such as lymphokine secretion specific to HIV antigens.
CC Cross-strain protection is achieved without use of adjuvant and the
CC synthetic DNA may provide a longer-lasting response. The env gene
CC includes a high proportion of codons rarely used in mammals, so
CC optimisation will allow this gene to be expressed efficiently in absence
CC of rev. The DNA sequences, more generally, may express antigens from many
CC other pathogens, e.g. hepatitis or tuberculosis, and also from tumours.
XX
SQ Sequence 269 BP; 72 A; 66 C; 72 G; 59 T; 0 other:
AAV21768 Length: 269 March 5, 2002 14:19 Type: N Check: 305 ..
Found using 'seq2-3' (pappu403.key)
1 GCGCCGGCATATCTAGACCACTCCCTGCGAGCTAAGCTGAGACCAATGACGGGTA
6 11
61 A
...
138 ATAAAAATGATCATCTCCAACTAAGACAGCGCGAGCTTCCGAGGATTTGCTCTGTT
188 193
198 TTATATATATTAAAGGCTGACCTGTCCGAGCCGCTGCTGCCGATGATGCTTGGA
258 TATCGCCCGGGC
259 264

23 matches found in sequence:
aav21771 ; VIR vaccination vector nucleic acid sequence:
(from "mycobacterieng.seq")
FOIG of: aav21771 check: 1773 from: 1 to: 3547
ID AAV21771 standard; cDNA; 3547 BP.
XX
AC AAV21771;
XX
DT 20-JUL-1998 (first entry)
XX
DE VIR vaccination vector nucleic acid sequence.
XX
KW VIR; HIV vaccine; codon usage optimisation; HIV; immune response;
KW human immunodeficiency virus; hepatitis; tuberculosis; tumour;
KW env protein; recombinant; vaccination; ds.
XX
OS Synthetic.
XX
PN MO9748370-A2.
XX
PD 24-DEC-1997.
XX
PF 17-JUN-1997; 97MO-US10517.
XX
PR 16-JUL-1996; 96GB-0014943.
PR 21-JUN-1996; 96US-0020165.
PR 21-JUN-1996; 96US-0020166.
PR 16-JUL-1996; 96GB-0014942.
XX
PA (MERI) MERCK & CO INC.
XX
PI Davies ME, Freed DC, Liu MA, Perry HC, Shiver JW;
XX
DR WPI; 1998-062825/06.
XX
PT DNA encoding protein with codon usage optimised for intended host
PT cell - specifically for DNA vaccines against human immune deficiency
PT virus, allows rev-independent expression of HIV genes
XX
PS Example 18; Pages 90-93; 113pp; English.

XX
CC This VIR vaccination vector is a derivative of the V10n vector containing
CC a new synthetic DNA, encoding a protein or peptide. The VIR vector was
CC constructed to obtain a minimum sized vaccine vector. The synthetic DNA
CC sequence encodes HIV env protein or a fragment and has codons optimised
CC for expression in a non-homologous host. The synthetic DNA sequences are
CC used to increase production of recombinant protein and to induce anti-HIV
CC neutralising antibody. HIV-specific T-cell immune responses or protective
CC immune responses in vertebrates, specifically as HIV vaccines. DNA in the
CC optimisation of codons results in increased expression of the DNA in the
CC host. The DNA induces antigen-presenting cells to stimulate cytotoxic and
CC helper T-cells, and effector functions such as lymphokine secretion
CC specific to HIV antigens. Cross-strain protection is achieved without use
CC of adjuvant and the synthetic DNA may provide a longer-lasting response.
CC The env gene includes a high proportion of codons rarely used in mammals,
CC so optimisation will allow this gene to be expressed efficiently in
CC absence of rev. The DNA sequences, more generally, may express antigens
CC from many other pathogens, e.g. hepatitis or tuberculosis, and also from
CC tumours.
XX
SQ Sequence 3547 BP; 869 A; 907 C; 839 G; 932 T; 0 other:

AAV21771 Length: 3547 March 5, 2002 14:19 Type: N Check: 1773 ..
Found using 'seq2-3' (pappu403.key)

158 AGTTCGGCGTTACATACTTAGCGTAATGCCCCGCTGCTAGCCGCCCAACGACCCCC

218 GCCCATTTGAGCTCAATATATAGCTATGCTCCCATAGTAACCAATAGGACTTTCCATTT
225 230
278 GACGTCATAGTGGTGAATATTACGTAATGCCCCACTTGCGAGTACATCAAGTATAC
278 283
338 ATATGCCAAGTACGCCGCCCTATTGACGTCATGACGTAATAGCGCCGCTGCATTATAG
361 366
398 CCCAGTACATGACCTTATGAGACTTTCCTACTTGGCAGTACATCTACGTTATGATCG
454
458 CTATTACCATGATGATGCGGTTTGGCAGTACATCAATGAGCGGTAGCGGTTGACT
459
518 CACGGGATTTCCAAAGTCTCCACCCCAATTGACGTCATAGGAGTTGTTTGGCACCAAA
547 552
578 ATCAACGGAGACTTCCAAAATGTCG
...
641 GTGTACGGTGGAGAGTCTATATATAAGCAGAGCTGCTTATGTAACCGTCAGATCGCCGGA
691 696
701 GACGCGATCCACGCGTGTTTGACCTCCATAGAAAGACCCGGAGCCGATCCAGCCTC
701 706
...
927 TAGCCTATAGGTGTGGTTATTGACCATTTATTGACCACTCCCTATTGTTGACGATACCTT
977 982

```

987      TCCATTACTAATCCATAACATGGCTCTTTGCCACAACCTCTTTAT
...
1220      TCTCCGGTAGCGGCGAGCTTCTACATCCGAGCCCTGCTCCCATCCCTCCAGCCGACTCAT
            |-----|
            1270 1275
1280      GGTGGCTGGGACGCTCTTGGCTCCTAACAGTAGGAGGCCAGACTTAGGCACACGACA
            |-----|
            1281 1286
...
1903      GTGCTGAAGAATTGACCCGGTTCTCTGACCCGTAATAAAGGCCGCGTGGCTGGCTTTTC
            |-----|
            1953 1958
1963      CATAGGCTCCGCCCCCTGACGAGCATCAAAAATGACGCTCAAGTCAGAGTGGCGA
            |-----|
            2000 2005
2023      AACCCGACAGAGCATTAAGATACCAGCGCTTCCCTCGGAGAGCTCCCTGCGGCTCT
            |-----|
            2049 2054
2083      CCTGTCCGACCTGCGGCTTACCGGATACCTGTCCGCTTTCCTTCGGGAAGCGTG
            |
            2142
2143      GCGCTTCTCAATGCTCAGCGTGTAGGATCTCAGTTGCGTAGAGTCGTCGCTCCAG
            |-----|
            2188 2193
2203      CTGGGCTGTGTGACGAACCCCGTTCAGCCCGCTGCGCTTATCCGGTAACAT
            |
            2261
2263      CGTCTTAGTCCAAACCCGGTAAGACAGCACTTATGCCACTGGCAGACCATGTAAAC
            |-----|
            2295 2300
2323      AGGATTAGCAGAGCGAGGTATGTAGGCG
...
2804      AATTTCCCTCGTCAAAAATAGTTATCAAGTAGAATAATCACCATGATGAGCAGACGAA
            |-----|
            2854 2859
2864      TCCGGTAGAATGGCAAAACCTTATGATTTCTTCCAGACTTGT
...
2954      ACCAAACCGTTATTCATTCGTGATGGGCTGAGCGAGCAAAATACGCAATGGCTGTGA
            |-----|
            3004 3009
3014      AAAGGACAATTACAAACAGGAATCGAATGCAACCGCGCAGAGAACA
...
3203      AATTCGTCAGCAGATTAGTCTTGACCATCTCATCTGTGAATCATTGGCAGAGCTAACCT
            |-----|
            3253 3258
3263      TTGCCATGTTTCAGAACAACCTGTGGCGCATCGGCTTCCCATACA
...
3360      ACCCATATTAATTCAGCATCCATGTTGGAATTTAATCGCGGCTCGAGCAAGAGCTTCCG
            |-----|
            3410 3415

```

```

3420      GTTGAATATGGCTCAATACACCCCTGTATTAAGTTATGTAAAC
...
-----
1 match found in sequence:
aav30037 : Oligonucleotide for synthesis of immunomodulatory cofactor GM-CSF.
(from "mycobacterieng.seq")
TOIG of: aav30037 check: 8235 from: 1 to: 61
ID      AAV30037 standard; DNA; 61 BP.
AC      AAV30037;
AC      13-AUG-1998 (first entry)
DT      DE
XX      Oligonucleotide for synthesis of immunomodulatory cofactor GM-CSF.
XX      Immunomodulatory cofactor GM-CSF; treatment; intracellular infection;
KW      immunogenic portion; antigen; intracellular pathogen; mammal;
KW      bacterial infection; legionella; tuberculosis; chlamydia;
KW      parasitic infection; rickettsia; leishmaniasis; malaria; viral infection;
KW      Herpes; HIV; FIV; ss.
XX      Synthetic.
OS      Homo sapiens.
XX      Key
XX      modified_base 1..61
XX      FT      /tag-
XX      FT      /note= "phosphorylated"
XX      PN      WO9812332-A1.
XX      PD      26-MAR-1998.
XX      PF      16-SEP-1997; 97WO-US16453.
XX      PR      17-SEP-1996; 96US-0025267.
XX      PA      (CHIR ) CHIRON CORP.
XX      PA      (SCRI ) SCRIPPS RES INST.
XX      PI      Lee WTL, Millich DR, Saliberg M;
XX      DR      WPL; 1998-212720/19.
XX      PT      Vector construct directing expression of intracellular pathogenic
XX      PT      antigen - useful for, e.g. treatment of intracellular diseases in
XX      PT      animals such as tuberculosis and chlamydia
XX      PS      Example 2; Page 40; 141pp; English.
XX      CC      Oligonucleotides AAV30035-43 and AAV30053 were used in the synthesis of
XX      CC      human immunomodulatory cofactor GM-CSF. The oligonucleotides are
XX      CC      overlapping, they were ligated together, and the full length product
XX      CC      amplified using primers AAV30035 and AAV30042. The amplified product is
XX      CC      cloned and used to exemplify the invention, which describes a method for
XX      CC      treating intracellular infections of warm-blooded mammals. This
XX      CC      comprises administering to the mammal a vector construct which directs
XX      CC      the expression of at least one immunogenic portion of an antigen derived
XX      CC      from an intracellular pathogen (e.g. HBV), and also administering a
XX      CC      protein which comprises the immunogenic portion of the antigen. The
XX      CC      composition is used to treat intracellular infections within
XX      CC      warm-blooded animals e.g. bacterial infections such as legionella,
XX      CC      tuberculosis and chlamydia, parasitic infections such as rickettsia,
XX      CC      leishmaniasis or malaria and viral infections like Hepatitis, Herpes, HIV
XX      CC      and FIV.
XX      SO      Sequence 61 BP; 14 A; 17 C; 19 G; 11 T; 0 other;
AAV30037 Length: 61 March 5, 2002 14:19 Type: N Check: 8235 ..

```

Found using 'seq2-3' (pappu403.key)

1 GATGCGCATCCAGAGGCGCGCTCTCTGTAACCTGAGTACAGACACTGCTGCTGAGAT
21 26

61 G

692 matches found in sequence:

aav35000 : Mycobacteriophage D29 DNA.

(from "mycobacteriophag.seq")

TOIG of: aav35000 check: 4492 from: 1 to: 49272

ID AAV35000 standard: DNA; 49272 BP.

XX AAV35000;

XX 27-AUG-1998 (first entry)

XX Mycobacteriophage D29 DNA.

XX Shuttle plasmid; vector: mycobacteriophage; lambda phage cosmid; antigen;

KW anti-tumour agent; lymphokine; immune response; tolerance; mutant; toxin;

KW autoimmune disease; antisense; vaccine; humoral response; T-cell memory;

KW cellular response; mucosal response; diagnostic; antibody;

KW infectious agent; ds.

XX Mycobacteriophage D29.

OS US5773267-A.

XX 30-JUN-1998.

XX 07-MAR-1996; 96US-0614770.

XX 07-MAR-1996; 96US-0614770.

XX 07-MAR-1996; 96US-0614770.

XX 07-FEB-1992; 92US-0833431.

XX 29-APR-1993; 93US-0057531.

XX 23-MAY-1994; 94US-0247901.

XX (UVP1-) UNIT PITTSBURGH.

PA (YESH) UNIT YESHIVA EINSTEIN COLLEGE.

XX Hatfull GF, Jacobs WR;

PI WPI; 1998-387007/33.

XX Nucleic acid encoding the D29 mycobacteriophage and related shuttle

PT plasmids - for producing recombinant mycobacteria and auxotrophic

PT mutants to express antigens and proteins providing long lasting T

PT cell memory

PS Claim 1: Column 7-50; 28pp; English.

XX This sequence encodes the D29 mycobacteriophage which is used in the

CC construction of a new shuttle plasmid containing D29 with an E. coli

CC lambda phage cosmid inserted in a non-essential region. This plasmid is

CC used to insert nucleic acid into mycobacteria by (non-)homologous or

CC site-specific recombination, particularly for expressing antigens,

CC anti-tumour agents, lymphokines etc., e.g. for evoking an immune response

CC or to induce tolerance to autoimmune diseases such as rheumatoid

CC arthritis. Such plasmids that contain a transposon or recombinant DNA can

CC be used to generate libraries of mutant mycobacteria, e.g. where

CC expression of recombinant mycobacterial DNA (to produce an antisense

CC molecule) will inactivate a selected mycobacterial gene. These mutants

CC are useful for selection of vaccine strains for inducing humoral,

CC cellular and/or mucosal responses and for expressing antigens for use

CC as vaccines or diagnostic agents, also other useful proteins such as

CC enzymes. Also shuttle plasmids specific for a particular Mycobacterium

CC can be used diagnostically, i.e. they are ligated to a reporter gene,

CC under control of a strong mycobacterial promoter, then incubated with

CC test sample and then examining for expression of the reporter. Vaccines

CC based on the new mutants prime long-lasting T cell memory, stimulating
CC a secondary antibody response that neutralises infectious agents or
CC toxins.

XX Sequence 49272 BP; 8912 A; 15900 C; 15401 G; 9059 T; 0 other;

50 AAV35000 Length: 49272 March 5, 2002 14:18 Type: N Check: 4492 ..

Found using 'seq2-3' (pappu403.key)

11 GGGGCGGACCATCTCTAGGAGTTTCCAAGCGATCAGAGACGCGGCGGCTTTC

61 66 69

71 CGCTCGCCGAGGCGCTGTAGCCACAGAGGCCAGAACCTGCGGACCGGAGTTTA

74 82 87

131 TACCCCC

232 ATGAGGCGCTGAAGGCCCTCTCTAGAGCGCTTACAGCGGCTGCTAAGAGGCGCTTTC

282 287

292 GGGCGCTCTGAGTAATAACCGGCCACAGAGGCCGGGGAATGTATCGGCGCAACCGCGG

293 298

352 GATCTAGCGCGCGGCTGAAAGCGGCTTATTGAAGGGTGACGCAACCGTATGCGGCAC

357 362

412 TCGTTCGAGTGGTACTGAGACACTCAACCGGGGAAGTTGACGTTCTCAACCTTGAGAT

452 457

472 GACGTTCCGACGCTCATCCGCTAGAGATCCCGATCTGGCGGCGACAACTTACG

472 477

561 GACATCGCAATACGCCGCGTTTCGGGCGGCGACTGCATCTTCTTGAGCGTATACC

611 616

621 GATTGAGACCGTCTGTGAGAGCGCGGCTTCTCCCGCGTGAAGCGGTTGAG

648 653

681 CTGCTCTGACCGCCGATTTTCAGCCCTCTGGGGGACATGCGACGGCGGCGGCTTGA

688 693

741 AGCTCTACCGCAGCCGCTGTGGCGGCGTACTGCGATCGCAAGGATGAGTGTCTCC

745

801 CAACCGCGTCTGGGACCCAGACAGCTTGACATTCTGCTTGCAGCGGATCCGAGAG

814 819

861 GGGCCAGGCTCGGATCAGCTGATGGCATCCGCTCTCGAAGATCGAACCGCTTTC

913 918

921 TCGCGCTGCTCGCAGAACTCTTCGATCGCAAGCACCGGAGCTGCTGGCTTACG

981 GCGGCGTGGTACTCGACGACATGACCTTCCGAGGTCAAGAGTACCGGACTACT

```

1041      |-----|
      GGGACGACGACGAAAGAGATATCCGACGATGGCGGAAAGAGATCCGCCGTAGCGG
      1047 1052
1101      CG
      ...
1249      GCGCGGCGACGCGAGCTGGCGCGGGTTCCGTGTTTCGACCGGACGCGACGCGACAGC
      1299 1304
1309      TGGACTTACAACAGCCCGATGGACCCCGCGCAGAAAGCGTGGGGTTTCACGCGCGCTGGCG
      1350 1355
      |-----|
1369      |-----|
      GTCCGCCCGCGACGACACTCACCGTCGGATGCCAAGCGCATAGCGAAGAGGAGACCAG
      1369 1374
1429      ATGCTCGGATACGTTTCGAGGGCGCTGGAGACAGTTGGAGGACGACGACTCCACCCGCTAC
      1468 1473
      1471 1476
1489      GACACCCGTAAGCAGCGGATGGACGCGCCCATGTCCAAGATCCTGAGCCAGGCGCAGGTG
      1503 1508
      1515
      1513 1518
1549      GTCCACCGCGACGAGCACTT
      ...
1580      TGGACTTGGGGCGGATCTCCTCGCTGTTTCAGAGTCCGCTTCTCGGCCGCGGACACGCC
      1630 1635
1640      CCGGTACGCGCGCTGTGTACGAGAGCGGTGTTCCGCAAGGCGCACGTCGCCCGGCTT
      1720 1725
1700      ACTCCGCGGCTGAGTTTCGACGCGGTCAAGCCCAAGGCGATGACATGTACGCGACACCA
      1720 1725
1760      AGCTTACTCTCAAGCCCGAGGTGCTGACCGGGTCACTGTGACCANTCGGCGACTCGCTGA
      1807 1812
1820      TGTGAGCGACAGCGTGTTCGCCGAAAGCGGGTGACGGCCTCGGCGTGGCGCGAATC
      1825 1830
1880      C
      ...
1885      GCGATCAGAACTCTGTAGACCCGAAATAGTCTCGGAAAGAAACATGCGAGGCGATCTAC
      1935 1940
1945      GAGAGCTTCAAGAAATAGCCCGAATGAAATTTTCATGCAATCGCAAAATTCAGCAGCGCGTT
      1999 2004
2005      ATTCTGAAAGACATCGAAGGTTGTATTACTACGCCGCCAGCTCTTT
      ...
      |-----|

2058      GACTATAACTGACAAAGCTCGAATACCGTTGGAGGTGGAATCTTAATGGCGCTGATG
      2108 2113
2118      |-----|
      CAAGCGACACACACCATTCGAGGGCTTTCTCGCGGTTCGAAACGCCACCGAGGCGCTCGTG
      2120 2125
2178      GCCGAAATGGCCACGATGATCACCCGCTCTCGGCCAACCAAGTGGG
      ...
2230      GGGAGGCGCTGGAGATCTTGTAGTACTCGGTTGACGCGCCAGGTGAGGTCAAGCAGAGC
      2280 2285
2290      AGTTGGCTGAGCGGAACACGCGCAAGCCAGATCGAAGCACAGATCATCGCGGAGGCGCGG
      2373 2378
2350      CAGAGTGAATTGGCGCGGCTCGAAGCGTGGCGCAAGAACTACCGGAGGACTGGAGCTGAA
      2373 2378
2410      TTACCGGCTCCCGGTCCTT
      ...
2460      GCCCGGCTGGCTCCGGGCTGCCACCGAGTAGACACCAAGAACGACGAGGACGACACT
      2510 2515
2520      CCGGCTTAACTTCAGCGGATTCGCGGTGTCTGTCGACCGGCAAGAGTCAAGCCGCTGAGG
      2536 2541
2580      GCGTAGCCCGACGCGCGGAACCTCAAGCCCGGAGAGCGACACGAAACGAAACGACCGCTG
      2616 2621
      2628 2633
2640      |-----|
      GCGCTCGTTAAGCGCGGCCAGGTGCTCTCCACCCAGAGGTGACTGTGGCACACAGTG
      2640 2645
      2643 2648
      ...
2710      AAGCGGACGAGAAACGCTTTCGTCGCAACACCCCGGAGAACCCGACCGAAACGATCTCG
      2760 2765
2770      ATGATCGGAGCGGTGAGATCCCGGAATCGGCGACATGAGTACATGGCGGAGACCAT
      2800 2805
2830      CCGCTCATCGAAGAGATGTACGAGC
      ...
2981      TGAAGCTGACCGGATCAACACAGATGCTCTCCGACTGTGTGACCGAAGGACCGGCG
      3031 3036
3041      GACGGTCCGCGCTGAGATCGAGCGGCGACCCCGACCCGACCGCGGAGGAGGCTGTTG
      3043
3101      ACGTAGCAGAGTGTCAAGCAGCGCTGCGCAAGGCGAGCGCGGGGTGATGCTCC
      3122 3127
3161      CCTGGCGGGGTTCTGAG
      ...
      |-----|
```

3310 CGATCCCCACCTGACCCGGTCCACCAGGCCAACCCCTGGCCCTGCGGGACCTGT |-----|
3360 3365
3370 TCCTCGAAGTGGAGACCGGGGGGAGACACAACGGCTGCACAGGT
...
3439 GCACGTACACCCGTGAACCTCAACGGGACCGACACCCCGGCATCGACTACAGAGCGTGT |-----|
3489 3494
3499 CCGAGAACCCGAGGGTCTGGCCGGTGACATCCAGAGCCATCGACGGCGCTCGGG |-----|
3534 3539 3548 3553
3559 CCGGTAAAGGCGCTGTGTCATCCGGTCTCGCTGTCCCGCGGTGAGCGCTGAACCTCAACC |-----|
3602 3607
3619 TGACGCCCCGCAAGCCCGCTGACCGAGCATTTGGTCAACAGATCAACAGAGGCCACGACG |-----|
3621 3626
3679 -|
ACTTCTTCGATGCGTTTCGACCAAGCTCCTCGGGTTCGACGTGAGATGACGGTCAACGACA
3680
3739 CCCTGAACCTTCAGATCAAGGTGACCTCGCGGCGTTCGATGAGTGGCGGCTCGTCA |-----|
3769 3774 3789 3794 3792 3797
3799 CCTTCGGGGTGGCGGTGACCTCGACGGCGGTCAAGAACTTCTCAACGGCTTCTCGGGC
3859 TGATCGGGCGGGTGAACACCGTCAAGCTCGACTTCTACTGGAACCGGACCTACAGACATCG |-----|
3882 3887
3919 AGTTGTCGGCGAGCTGGCCGAGACGCGCGGTACCCCGCTCCACGGCGAAGCGCGCGGTC |-----|
3941 3946
3979 TCAAGGGACCTCGAAGG
...
4132 TGCTTCATCTGCCGACCGGGAAGCGCGCGGTGACCCCAACAACTGGGCTGCT |-----|
4182 4187 4185 4190
4192 ACCGGCAACCCCGGTAGCGCACCACTGACGTGAGCTCAATGGCAGAG
...
4388 CGACTCCTGGCGGTATGCCACCACCTTCCCGTTGCTAAATCGGTAGAGCGCTGGC |-----|
4438 4443
4448 TCTGGACAGGTAATTGAGGTGAGTCTTTGGCGGGAGACACCA
...
4591 GACTACAAAGCGGAGATACCCAGCGGCAACGGGATCTGACTGTAAATCAAGCGCTTGG |-----|
4641 4646
4651 CTTTCGAGGTTTCAAGTCTCTCTCGGACTTGACAGCCACCAAGAAAGAACCCATGAC |-----|
4708

4711 -|
GCTCATATGTCACAGCGCACCGCAGTGGGTCCAGACATGTGCCGCTTCGCGTGG
4713
4771 CAACAGGTACGGCTACGGCGGGGGGCTTCACACTCAACCCCGAGACACACCACTGCTC |-----|
4792 4797
4831 GGGTCTGTCTTCGAGA
...
4878 GGATCGGAACCGGTACGGCTCGACTGAGAGCTTCGGGCTGACCAAGACAGATGTCACG |-----|
4928 4933
4938 ACCTCGGGTTCAAGGCGACTCCCTCCGGGAGCGTTCGGCGCCCTGGATTCAACCCGGTCA |-----|
4950 4955 4967 4972
4998 TGCTCTGCGGGCTTCACACAGCGCGCGGGGGCGGTACTCGCACACCGCTTGCACGCTGA
5058 -|
TGACGATGACATTCGCCGGTGGCCCGGTGAAGTCTCGCAACGAGCGCTGAGTGGAGT |-----|
5059 5064 5102 5107
5118 CCGAGAGAAAGTCAACGGCGTGGGGTGTTCCTCTACAGCGGCGCACGGCGCTGGAAGC |-----|
5174
5178 -|
ACCGCTCTTCCACGACTTCTGTACTCTGACGGAGGAAGTTGAAGAGCGCCGACGACGA
5179
5238 GTGTGACGCTGCGGAAATCCTCGCTCGCGCAACGGGTCGCGTACACCGAGCGGTAG |-----|
5243 5248
5298 C
...
5343 CCAACCCGAATCGCATCGCATGTGGCTCGCCCAATCGGCATGAGTCAAGACATTTCA |-----|
5393 5398
5403 AGCCACTGGGAGTACGCCAGCGGGGAGCGCTACGACACCCGAACCGACCTCGGCAACA |-----|
5429 5434
5463 CCCCAGAGTTCGACGAGACGG
...
5614 GTCGAGACTGCGCTGGGCAAGCATCGGTGCCGCTGTACTGGACCGTCAAGCCCCAGA |-----|
5664 5669
5674 CATCAACGCACTGACGACCGCGCGCACTCGAANAAGGTACAGCGCGGATCAACGGGGG |-----|
5687 5692
5734 GCTTCACCAACCTCGATGACCGCGCGGTACCAACCTGGCCCTCGCTGTGGCGGACCA |-----|
5757 5762 5786 5791
5794 ACTACTACTCTGTGTCGAGATGACGAGCAATTGGCTGATCCAGATTCAGCGCTTCAT |-----|
5816 5821 5836 5841
5854 CCGGAGATTCACGGGGCGCTTTCACACCGTGTGACGCAATCCCCCTACGGCGACCC |-----|

| | | | |
|------|---|-------------------------------|-----------|
| 5914 | GCAGAACCCGAGCGCTCGGAGCCCGGAGCAACCTCTGGCAGCTCCATGAGCTGATCA | 5869 5874 | 5906 5911 |
| 5974 | GAACGGACGCGCATGGGGCAAGCCCGCTACGTCGAGGAAATCGCGGCAAGCCGTTGACCT | 5978 5983 | |
| 6034 | CCGCGAGCTGGAGCGATTGTCCGCGCCCAAGGAGCTTGTAGGAGTCCCTCCCGCA | | 6081 6086 |
| 6094 | GTTTCATCCGACGGCGTCCGAAAGTCTGTGCCCCAGATCGAGCGACCAACCCCGAGTACTT | | |
| 6154 | ACAGCGCTACATCGCCAGGAATGAGCCCTATGAGCCCAAGATCCGTGAACCGCTTAC | 6164 6169 | 6204 6209 |
| 6214 | TACGTCGCACTCTGTCTCCCGGCATCCCTGGGCATCCGCTGATCTGGGGGGGATCGAC | 6247 6252 | |
| 6274 | GCGGGCGCAGCGCGGAGACATCGGGGACATCGTCGCTGGCGCTTCACACTTGTGCGCGCA | 6295 6300 6309 6301 6306 6315 | |
| 6334 | GCCGACCGCGCCACGCGCGCTGTCAAGGTCAA | | |
| 6371 | AGCGCAAGATGACGACGCTGACACACTTCCCGGTGATGATCAGTCCAGGAGGCGCTCGAGC | 6421 6426 | |
| 6431 | AGGTGTCGCGGCGCAAGACGATGAGGCTGAGGTGAGCGCGGTCAAGCAGGCTCTG | 6451 6456 | |
| 6491 | AGTCCGCTGTCAACGG | | |
| 6724 | GCAACTACCCCGCTGCGGCTTCCCGATGTGGCCGTGCGTGAAGAGGTTGTCGCGAGC | 6774 6779 | |
| 6784 | TGATCTCTGAGATCGAGCTGAGCTGAGCGGAGCCCTACGCGGA | | |
| 6958 | ATCCCATGCGGCAAGAGGCTTTGCCCACTCTGACGAGTGGATCCACCCGCTGCTGCCC | 7008 7013 | |
| 7018 | CTGACACCCCTCGGAATCCTCGAGGACCGGCTCGAAACCTGAGACA | | |
| 7075 | AGTCCGCGACTACGCCACGACGCTGATGTAGTGAAGGCAAGCGGCTTCATCGTGGCC | 7125 7130 | |
| 7135 | ACGAATACGAGTGTGCGCGCGGATCGTGTGAAGGCAAGCGGCTTCATCGTGGCC | 7146 7151 | |
| 7195 | GGGACTCCGTGTAGCCACACTCATCGAGCTTGCCCAAGCGTCCGATCACCGAGGAATTG | 7231 7236 | |
| 7255 | CGTTGGCGGACCATCATCGACGCCCTCACGTTCTTCCGCCGCTCTGTATGGCGACA | 7275 7280 | 7308 7313 |
| 7315 | AGTGGCCGACCTCTACACCGCTACCCGGGGGTGAGTTCTCTACGACA | | |
| 7457 | CGGTGACGAGATTGGATCTGCCGATCCCAAGATGACCTTGATGGGCGCTTTGA | 7507 7512 | |
| 7517 | AATGGCTGTGAGTACGTCAATACCCCTGGGGACATGACGATCCGACCCGCTCAAGG | 7555 7560 | |
| 7577 | TTTGTATCTGCGCTGTCCGAAGCAGACTGCTTGA | | |
| 7765 | GCCGAACTGTGTGCCCAATAGCATTTCTCAACATCTGATAGAACCGGTCAGGGATCGC | 7815 7820 | |
| 7825 | AAGCCGCGCCCGCAGCGTGAATCACCGTCCGCGCGCTCAGCCAGG | | |
| 7981 | CGCATCGAGGACGACACCTCTCTACCAAGCATTCGATGAGGAGGTAAACCGCCGACTTCTGTC | 8031 8036 | |
| 8041 | GTTTCAGAACGAGAACCCAGTGTGGGGCCAGGGGCTGACGGCAAGGCCACGAA | 8043 | |
| 8117 | TCGAAGCACAATGACCAAGGTGAGGGCTCTGCGAACCTGTGATCTGCAAGCCGACA | 8167 8172 | |
| 8177 | TCCCCGCGACCGAGACGCTGCCGAGAGAGCGTACGTGAGTGGACGACGTCCG | 8194 8199 | |
| 8237 | GGAATCTCTGACACAGCATATGTACGACGCTCTCGAAGCGCCGCGACACCCCGA | 8266 8271 8277 8282 | |
| 8297 | TCTCCGAGATCCCTCTGAGAAAGAGAAATCCGACGAGGTTCCGGAGGCGATCGAGAAGC | | |
| 8357 | TCCGCCAGGGGCTGTTGATGCCCGAGGCGACTTCACGTGGCTGCCGATTGACGACATCA | 8374 8379 8388 | 8407 8412 |
| 8417 | TCAAGTCACTCTGTGACCAACGAACTCATCATCAGATCAGCGCG | | |
| 8579 | CGCTGGACCGGGCGCAGAAAGATCACCTCGGGTTTCAGCGGGTCCAAAGTCCCAAGACTTGA | 8629 8634 | |
| 8639 | CGGCTGTGCTGCGGTGCCGGTCTCCGACGGCTGCTGTTCGTAT | | |
| 8687 | ACATCTGGGATCCCAAGATACGCGCGCAAGTCCCGCGAGTTCGTGGACGCTGCAG | 8737 8742 | |
| 8747 | TGCATTCCCGTTCTCCCGGTACGACGTGTGCGCTTCCGCGCGCA | | |

...
8807 AGGCATACGTGACGTCTCTGGGGCCGGACCTACAGAAGAAGCTCAAGGTCAAGGCCAGCC |-----|
8867 CGAACAACCCGGTTCGACATGCGCGTCCAGCAGAGAGCGATTGCGGTTCGACTGTG |-----|
8927 AGGCGCTGGAGAGACGGGTCCTCGAAGGCGAGGTCTGGCAGCAGCGCAACCCGGTCTTC 8880 8885 8906 8911
8987 GGCAGACGCTCTGAACGCCAAGACACCCAACTACTACGACGCGCATTCGCATTTCGA |-----|
9047 AGGTCAACCAAGACTCCAGCAAGAAGATCGACGCTGAGTCTGCGCTGTCTCGGTTTCG 9001 9006 9013 9028 9033
9107 GGGCGAGACAGACTACTCATGAG |-----|
9148 GGCCGGGTGTGGCCGTCCGATGACAGCCCCGCTCCCCGGACAGAGAGAGATGCTGATC |-----|
9208 CGGCGATTCGCCCGAGACGAGATGTCTCGGCGTTGAGGACAGCAACCAAGACCTCCGAT |-----|
9328 TCCAGATGCAATGCTGCTGCTCAGCTCGGATACCCCGGCTGTACTGCTGACTCATCG |-----|
9388 CAGAGCGAGCGGCTGTGAGGGTTTCCGGCTCGGGACGCGCAGCAGAGGGCGAGAGAGC |-----|
9448 TGTGGCAGTGTGGCAGAGGCGAACAACCTCGACATCGAGGCCCGCTGGCTACACCGAGC |-----|
9508 CTTACGTCCACGCGCGGTCGTACATCAGCATCAGCCGCCCTGACCCCGACGATGACCTTG |-----|
9568 GGTGGGATCCGAACGTCCTCCGCTGATCCGGGTGGAGGCCCGGACGCGGATGTACGCGAGA |-----|
9628 TCGATCC 9579 9584
...
9675 GACGCAAGAGGCAACGATCCAGGCTGCCACGCTTACACCCCAACGAGAGCTTGGG |-----|
9735 TGGTTCCGGGCCAAGGCGAGTGGTGAAGTGGTTAGCGACCCCAACGAGGCTCGGCGCG |-----|
9795 GTCCCGGTGGTCCGCTTCCGAACGAGACCGG 9771 9776
...

9864 GAGCTTGCTGATGACCCAGCGCGGCGCTGCAATCTGATGCTGATGCGAGGAGACTGCG |-----|
9924 GAGCTGATGGGCGCTGCCCCAGCGACTGATCTTGGCATCAAGCCCAAGAGATCGGCGTA |-----|
9984 GACCCCGAGACCGGACGACGCTGTTCGACGCGTACTCGCCCGCATCTGCGGTTTCGAG |-----|
10044 GAGCCCGAAGCAAGATCCAGCAGTTCTCGGCAAGCCGAGCTGGGCCAACTTCACCAACGCA |-----|
10104 CTCGATCAGATCCGCCAAMACAGGTGCGCTACACGGGACTCCCTCCCACTGAGT |-----|
10164 ACCGCGGCTGACAATCC 10113 10125
...
10221 ATCAAGAAGTCCGAGCGGAAGAACGGCATCTTGGGCGGTGCTGGGAAGGCGGATCGC |-----|
10281 CTGGCCATACCGGCTGATGAAGGGTGGGAGCTTCCCGGACATCTCCGATGAGAGCG |-----|
10341 GTCTGGGTTGATCCCGAGACCCCGAGCTACGCCGGAAGGCTGACGCCCGCAGAAAGCTC |-----|
10401 TATGGCAACGGCCAGGGCGTGATCCGCGTGAAGCGGGCTCGCAAGGACATGGGCTACTCC |-----|
10461 ATGCGCGAGGCGGAGGAGATGCGCGCGCTGGAGCAGAGAAAGGCCGGATGGGCTCGGC |-----|
10521 CTGCTGGGCAAGATGTTGACGCGGACCCGACGCTGCCAGGCTCCCAACCCCAACGCCCA |-----|
10581 GCTCCCAAGCCACAACCGGCGCATCGAAAGGGGTGATTCGCGCTGACGCGCTGAGCAGTATG |-----|
10641 CAGCGGCCAGCGCGTGATCACTGCGGGGCTCGCGGGT 10624
...
10720 CTCATTCGCGAGTGGCTGGGCTACTGCAAGTGTGTTCCAGAGGTCAGGCGTGTA |-----|
10780 TGGGAAAGTGGCGACCTGGCGCGGACTTCTAGCACTCCAGCGCAGACTCCACACGCC |-----|
10840 GGAAGTTCCCGCAGAGAGAGGTTGGGAGCGACCTCCAGTGGAGTGTTCGTCAAGAA |-----|
10900 CATGAGCCCGCAGCAAAAGGGTTTTCGACAGGCGGACTCTCTCAAGCTCGGCTCAACAA |-----|
10960 GCTGACCTTGGCGCAGATGCGCAGAGTGAAGTGGCAGGTCGCGCAGATCATTCGCGCG |-----|
10969 10998 11015

```
11020 TGTCAAGACGACCCAGCTCTTAAGATGCTGTAAGGGCTGGGGGAGGGTCGCGACACCGGGGCG |-----|
11020 11027 11066
11080 CGAATCATGGCCCTGGTGTCTGATGCTGATTTTCCCGTGGCCC
...
11292 AGAACTGGCCCGGAAGGCCCGACAGAAAGCGGCTCTGACGCTTGATCGACGCTGGGA |-----|
11352 AAGAAGCCAGCCAGCTCATTTGATCTGSCAAGGCCCGCTCCAAAAAGAGAAACAGAGAGA |-----|
11412 CGATCAAGCTCTCCGTCGCGCTTGTATCGCGGAGTTGCAATGTCCGACTACGCAC |-----|
11415 11422 11432
11472 TCGCTGCCTAA
...
11520 GGAGGCGAAACACATGTCCGACACCGCACTACCGAAGGCACCTCCGGCGCGACCGGA |-----|
11580 CCCCCTGTGTCATCTGACAGCCGCTGGAACCGACTCCGAGAACCTTA
...
11667 CTGCTCGGGTTGCCAAGAAAGACGCACTGAAGCCGCAAGTCAAGGCGCGACGACGCTC |-----|
11717 11720
11727 ACACCGCTGAACCTGCTGCTCGCACACTGCAATCACCAGACTGGAGAA
...
11790 CTTGATCCGGCTGCAGAGCTGGAGACTGCTTGGCCGAAGTTCCGACGACCAAG |-----|
11850 TGGTCGGCTTCTGTGACATCTTGCAGGGCGAGAGCGCCGACTCGATCGCCGAGTCCGCGGA |-----|
11882 11894
11910 AGAAGAACCTCGAACTGATCGGGCGTTTCGACAAAGAACCC
...
11973 AGGGCTTCGGTGTGTCGGCAGGAAGTCCGCTCAACGCTGACCCGATCTCTAAGCCCATGA |-----|
12033 AGGGCGTTCTCGGCATCAAGTGAATGTCAGCTGAACCCCATTCCTTAGACAAGAGATTAG |-----|
12035
12093 CAGATGCCCGCAGGCACTGCTTTTCGAGTTGATACGCTCAGCTCGCCGACGAGCGGAT |-----|
12135 12147
12153 ACCATGTTCAAGGGCTACCTGAGACCGCAGACGGAAGAGACTACTTCCGCCGAGGCCGAG |-----|
12213 AAGACTCTGATGCTCGAGCAGTTTCGCCAAGAGTGCATGGGTACACAGGGCCAGAAAG |-----|
12222
```

```
12273 ATCCGCACTGGGTGGCGACCTGAATGCCACAGTGAATGTTGAGGGTGAATGAAGCCC |-----|
12288
12333 ATACCAAGAGGCAACATGACTTTCGAGACCATCGCGCCCAACAGATGCGCAGCATTTTC |-----|
12393 GTGGCTCTGGCGGAGACCGTCCGTGCAACCCCGCCAACTACTGTGGAAACCATGCGTACC |-----|
12395
12453 AAGTGGGAGACCGCCTTTCGCAATGGCGTTTCAGAGGCGGGCGATGACAGCGACAGC |-----|
12458 12476 12491
12513 CCGTTCGCGACCTACATCGGTGACACCAAGG
...
12574 CCAGACCGGTGTAGCAGCAGTGGCCGTGAACGGGCTGAGCCCTGCTGTGATGAACGAGCA |-----|
12634 AGAAGTGAACCCACACCCCTTCTGAGCAGATCACCGAGCCGATCTGAAGCGCGGAAGG |-----|
12694 ACCAAGACGTCGCCCGCTGTTTCATCGAATCGACCTACGTTGAGGCGCGAGCCGTTCC |-----|
12703
12754 GTTGGGCGCGATGCTGCGCCGTCGACACATCTACGACACACGTCGTGAGAGGGACACA |-----|
12765 12789
12814 CGGTGCGCTTCATGGGTGACTTCTCCAGCTCATCTGGGGCAGATCGCGGCTGTCTCT |-----|
12874 TCGACGTGACGATCAGGCGACCCCTGAACCTGGGACCGCTGAGAGCCCGAATTCGTCT |-----|
12934 CGCTGTGGCAGCACAACTCTGTCAGTCCGTGTGAGGGCTGAGTACGCTTCCACTGCA |-----|
12994 ACGAGCCCGAGGCGCTTCTGTCGCTCTGAACCAAGTGTGACGGCGGCGCGAGGGCTGAG |-----|
12998 13004 13011
13054 CCTAATTTGACAT
...
13130 CCAATCCACCGTCAAGGTTGGGTTTCGGGAGGTCTCCGACGAGTACGCCACCGACGCTGAT |-----|
13190 CCGGCGTGGCGGTTGGAAGCGTCTCGGAAGCTTCGACACCAACCAACCAACCCGCTCC |-----|
13250 GAAGCAGGAGCCCTGCGACCGAGGAGTAACACATGCTTACGCGACGCTGACGACGTTGT |-----|
13310 GACGTTGTGGCCAAAGGAGCTTGAAGCCAGAAATCATATGCGCTGATGAGGCGGCACTCGA |-----|
```



```
13310      13346      13357
13370      ACAGGTGACGCGATGATCCGGCTCGGATCCAGATCTGGACGCCAGGGTGTCTTGGA
13390      13410
13430      CATCTTCGGGGCGGATCTGATCGACATGAGGCCGA
...
13536      CTCGAGGCTGATCTGTCTCAGGGCAAGCTCGTATCTCTTGACGAGAAGTGGACGACCTTG
13586
13596      GGAGTCAACCGACTCTTCGGCATGTCCACCTCTGTCGCCGAACATCTGATGCCGACATGA
13655
13656      GCGCCAGCGGACGTTTCAGCGGGGCTCCGATCAAGTACCCGCGGGTTTCTCTGCGGCTCA
13660      13669
13661
13716      CACCTGACCAAGTCGACGCGCGATGTGGACCAAGAGGGATCTCCGGTCTGCTACT
13730
13776      GCGTCCACGATGGCGCATCGAGTTCGGCAACGTCCTCTCCACGCCCCAAGCCGAAAGCA
13805      13831
13836      CGTACATCTGATGAGCTCTCTCGATACCGGAGCCCGGTACCAAGCCGCTCT
13836
...
13909      ATGACCGCGGACGGGAACAAGAAAGACCAGCCGTCGAAAGCCCCGATCCAGAGCATGCA
13959
13969      CGGTTCCAGGTCGGCCACACAGTCCGATAGTCCGGCAGAGTGCTGTAGACGACAGGGG
13978
14029      GGGTT
...
14048      TCTACCGGATGCGGTTCCCTCGCTCTTCAACAAGACAGCGGATCTCTGGCGCTCAGA
14098
14108      CCCAGATCGAGTGAAGGGCCAGCGGTGGGGGCTCTTTGGAGACGCCACGAGTAGACT
14136      14149
14168      CATGCGCCCGCACTGGCGGGTGCATACAGATCAAGAGGTTCTGATGGCGAAGT
14169
...
14245      TCGGCGCCAGGACAGTGCATGTCGCAAGGGGTCAAGAGAGAGCGAGCGCTCACCC
14295
14305      GCGGTGCTCGAACCACCTGGCCAGAGCAACAAGAGACCCGTTATCAACCAAGAGGGGT
14339
14365      ACTTCCGGCATGATCGAAGAGGTGACGGCGATGTGACTTCCACACGGTCTCTGCACG
14394
14425      CGCCCAAGCGGTTGCCCTTGAGTT
...
14514      CTGACCCGAGCGCCCATCGGGGCGACCGCTCTCTAAGAGAGACATGGGGCGCATGCC
14564
14574      CGAGTACAAAGTGTGTTGCTCCGATCTCCGAGAGACCTTCGAC
...
14659      ACTTCCGTGAGTTCCTCGATGATCAACATCCCGCATCGCGGATCGAGACGCAACG
14709      14715
14719      CCCCAGAGCTGCACCTCGCTGCGGGTGTGAGATGTGGGCTACTCCACTGA
14720
...
14829      GTGAAGAAGCGMACAACAACCTCCCGAGGGTATTTGAGTTGATCTTGAAGCATGGG
14879      14886
14889      GCCACCGAGTTCAAGCTCCTCTTACAGAGACTCTGGGAATCCAGGGCTGATCAGGCTC
14891
14949      GCGGTCCGCACACCGAATCCACCACTTAACCGAAGGTAGCACATGGCAGAAAGCAC
14949      15003
14949      15006
15009      GATCGAGTTTGACTGCTGGGTTGGGTATGTATGTTACGTCGCCAAGACGACCGCTGCA
15045
15069      CATACGGCGGCGGAACGTGAAGACCATGACCTGTCCGACCCGTCGACCTGAGCTGGAACC
15011
15129      ACCGGCTGTGTCGAGCGTCCGGCCACACGACCGAGGACGCTCCCGAGTTGCGCTCGAG
15141
15189      GCGCGCATTTCCGAGTCAAGGGCTCTGTGGCAGAAGAAGCTTCCGAGATCACAC
15192
...
15249      GAGATCCGATCGACTACGTGCTGTCTGTCTGCACCAAGTTGATGAGCATGCTGCTGGC
15299
15309      CTGTACTACGGCCCGAAGCGCTCCACGACCCCGGTGTTCGGTGTGAAGACCGGCGAG
15324
15369      ACCAAGAGAAGGGGTCTGGTGTGATGAGAGACGGCGCATGCGCTTGGCCACAC
15405
```



```
17825 GTGACAGCTTCAGCAGCTCTCCGAACGCTGGTCACCTCGCTCGGCCGCTACTGCCC      |-----|
17851
17885 CAGATCGGGAGACGTTTCGGCCAGATCTAGTGGCTCATTCAGTTGGCCCCGAGATC      |-----|
17896
17945 ATCTCGCTGATTCGCCAGCGTTCCAGCGCTGATCCCGCGATTCGACAGCTCGCTCCG      |-----|
17950      |-----|      |-----|      |-----|
17962      |-----|      |-----|      |-----|
17971      |-----|      |-----|      |-----|
18004
18005 TCGCTGGTTCAGATCGTCCAGCGCTTACCACCAAGCTGATCGCGGTCACTCGCGGTGTC      |-----|
18009      |-----|      |-----|      |-----|
18017      |-----|      |-----|      |-----|
18025
18065 CAGATGCTATCAACCTTGCGCGCGCTGGTCAGCGCTGGCGGCTCCATCGCGCTGTC      |-----|
18068
18125 CTGATCGGTGGCATCTCCGGGCTGGTTGGCTCTGCGACACTGCGTGGCGGAGTGGC      |-----|
18132
18185 GAGTGGGTGGCTCTGTGTCCAGCGGTGTGCAAGTGTCCGACTTCGT      |-----|
18192
...
18255 TCAAGAGCTGTTGATGACGACGAGCTCTGGCTGATCGAGCGGGCAAGAACGTCGC      |-----|
18305
18315 AGGCTCTGATCAACGCGATCGGTTGATGATCAGCTCTCGGGTGAGCAAGGCCAAGAAC      |-----|
18330
18375 TGGCCAGCAGCGTGAAGAGCGTGTACCGGCTTCTCCGGGATCCACTCTCCGTCAGAGG      |-----|
18392
18435 TGTTCGCTGAGAT      |-----|
18442
...
18494 TTCCACCCCGTCAATCGAAGAACGCCAAGCCCTGGCGGCTGAGCTGTCTCAGGGGATGAG      |-----|
18544
18554 TCTGGCGTGGACCCCTCCGGGATCTCGCTGTATCAGCACCAAG      |-----|
18559
...
18644 ATCCAGGTGAGAAAGCAATTTCCAAAGAGACAGGCCGCCGCTGGCGGCTGCAG      |-----|
18694
18704 GCGCAGCTGCACAGATCAAGGCGCAGAAAGACATCTCGCGTTACCAAGAGGACCCGATC      |-----|
18711
18764 AAGAACGAGAGACTACGTGCGGCGACGCGCGATGACCCGCTCGTAAGAGCGGCTTCC      |-----|
18794
18824 GCGTTGATGAACGACCGGTCGACTTCGGCAAGCGACTGGGAAGCAGTTCTTTACGGAC      |-----|
18856
18884 CTGGGTATCTCCGGGAGCGGTGCGATCT      |-----|
18891
...
18917 GCCATCACGAGGGGATTCAGTACATCTTCAGATCGGCTCTGTGATGAGCGCGCTGCG      |-----|
18967
```

```
18977 ATCAAGGACCGCGAGGAAATCAAGAAAGCGCTGTGGTGTGGCGGAGCTGTGACATCA      |-----|
19037 CCAAGAGTATACCATTTGATCACCAGACCACTCTTGAAGTGTCAAGGTGAG      |-----|
19066
19097 GCTTCAACTTGAAGACCGCGGTGACCAAGGCGGTGTTCCGTGCCACAGAGGTGAGGTTGTT      |-----|
19099      |-----|
19107
19157 TCTACGACCTCCCGTCAGAGTGTCTACGAGAGAGCGGGGAACTACCCCGGTCTCGCT      |-----|
19177
19217 ACCTGGACACGAGTTCTGAAGCGGACATCGTCTTCCGGGTTCAAGATCTCAACGACG      |-----|
19246
19277 CGAAGCAGGAGACCGGCTCTGGCTGTGCGGAGACTCCGTGTGGCTAAGCATGCGCGT      |-----|
19332
19337 TCAACCGCGTCTCAAGCTCTACGTACACACCCCGGACTCCGGTACCCTGTATACC      |-----|
19337
19397 TGGCGCTGTTGAGTCCCGCCAGGTGAGATGAAGACCGACCGCGGTGCAACACA      |-----|
19398
19454 TCAACTGAGGTTGATGTCTGTGATCTCTGTACGACCCCGTTCTGTGACGAGAGAGCGAG      |-----|
19504
19514 TGTTCGTGATCAAGACCAAGATCCGATACAGTTTGCACCCGAATCTGTGACACCGCGT      |-----|
19519
19574 GGCCTGTGGAGAACTGCCCAAGAGAGCGTGGGATCAAGTGTGCGCGGAGCAGGCGG      |-----|
19599
19634 GGCTCAACCCCGACGACAGT      |-----|
19641
...
19771 ATCCCGGACTACTGTTCGAGAGACGAGAGTTGCCAACCGCGGCTCAAGACCGGGG      |-----|
19821
19831 TTGATCTAGGCGAGAACTGCATCATGACACGACCGACGCGGAGG      |-----|
19836
...
20023 CTCGCCGAGCGGTGTCGCTGTGGGCTGTGGAATGATGGGCTGACGAGGTTCCCG      |-----|
20073
20083 AGGCTGAAGATCTGTGCGGAGAGATCCAACTGGCGGCTCAAGCGGAGCAGAGAGGAC      |-----|
20116
20137
20143 TGAAGCCACCGGACGTAGAGCTGCGGACCGGCACTTCCGTTCGCGGCGCTGTCGCGG      |-----|
20172
20203 GTAGCGACTGTGTGAGTGGAGTTTATCGAAGAACGAGACCGGGGTAGCCACGCTGACG      |-----|
20206
...
20272 GCCACTACTGTGGCAAGTGGGTGATGAACCGGCGGTGACGCAAAAGCGCAACGTATCC      |-----|
18967
```

```
20332      TCAAGCTGAGAGAGAGCGCTCCGATGAGCGGAGTATGAGCACAATACCGGGTGTCA      20322
20334      20349
20392      AGGAGACTCCGG
...
20529      GCCAAGTGGTCTTGTCTGTACGCTGTTCTGTAACCTGTCTGCACTGGAACGTCCTG      20579
20589      TGGAGATCCCCGAGCATCCGACGAGACATCTGGAGTGTAGGGCCGAGCTTCAACCC      20591
20591      20601
20649      AGCAAAATG
...
20679      CCTTCTGTGTGACAACTCGCCCATCAAGATGATGTGTACGCGGTTGCGGACGTTTAC      20729
20739      GACAGCGGGAAGCAGATCTTGTAGAACCAACGACTCAGCTGACGCTGTGCGCGGTACATC      20785
20799      AAGGACCGGACCGGACCCGTTGATGATCACTGAAGGCTTG
...
20932      GCACCGAGACCGCTTCGGCGGCTCTGTGCTCACCGGGTTCGCGGCGGTGTCACAC      20982
20992      TGGCTGAGAGACGGCCAGGTGCAAGGGGTCCGACGATTCACGGGTGA
...
21087      CCGAGGGCTCCGACAGTCTGTTCGAAGAGGGCCACTGACCGGGATCAAGTCTCGGAG      21137
21147      TTCTCTACTACAGAGCAACCGACCAAGCTTCTTGGCCGGTGGAC
21193      AGTCCGCTCCGGGATCAACGAGGGGATTCGGCCCTGTGTAACATGGTGGCGATCTGC      21243
21253      TGACCTCTGTTCAATCAACAGCCAGCTGCGCGTCTCGGCGGGCTCGG
...
21342      GTCTGAACCCCGCTTACTCTCCGAGGTGTTCGGCGCTTCAATGGAAGTCCGAGCGCTGCT      21392
21402      GCGATGGGATCTCGCTTCCCATTTGGGGGCTTGAGAGCAATCGTCAACCGGCTGGGAG      21441
21462      TTCCTACTACGAGAACATGTGTGAGAGCCGATGAAAGCGTTTCAACCTCTCTGGCGTTC      21500
21522      GCGGCATGCGACAGCATCCAAAGAACCCGAGCCGAAAGGCTCACA      21515
...
21758      CGGATACCGCGAACCGAAGAACCCAGCTCTGCACATCTCGAAGAGATCAAGCTTTCAA      21808
21818      GGGCGCTATGGGACAAAGCGGGATTTCTTAACCGAAAGGACAGCCGATGATTCGCTCAC      21819
21878      AAGAGACTCACAAATCCGAGACCCGCGACAGCAGCTCGTCTGGGCGTCCGCAATCTCC      21895
21938      CGTTGATTGCAGCGCTCGGGGATTCACGATCCGCGCTACCTGCGGATTTGTCAGAGC      21949
21998      ACTTGTGGAAGTCCG
...
22248      TCCGAGGCCAGCGGACGACATGAAACAACGACGAGGTGGGCGCGAAGAGCGCTCCCG      22298
22308      ACCGAAACCCGTGCGTATCCAGACATTTGCCAAGCTCACAGACGAGAGAACCAGGGA      22363
22368      TGCTCGCACAGTACGAAACGAGCGGTGTGATCAAGAACGACCGCCCGCCACGATGG      22403
22428      CCGAGTCTGTGAGTGAAGGAGCTATTCAACCCGATTAACCTTGGGA
...
22600      ATCACCCGAGGTTCAAAAGAGTCCGGAGGAGATCCGACATTCTATAGATGAGGCGGTAA      22650
22660      ATCGAGCGCGCGAAGCATGCTGTGTGTGACGAAAGAGTCTGCTTGCATTCCCAACCA      22664
22720      ACCCACTCGAAGCGATCGAGACTGACGCGCATTCGAGATCGCGGGTGTGATTTCACT      22730
22780      TCGGCCAAGACTACACCGAGCAGATCAATTGCGTGTCTGTTACAGATGCCGCTTCCGAC      22811
22840      TCGAACACGGGATACCCCTCTCCGGGAGGACCTGTGAAGCTCTCTGTGAGAGCTCTGC      22892
22900      AGAGTTCAAGAGATGATCCCGGACTGGGCCGAAGCGGGTTCGACA
...
23053      CCGGAGAGTTCAATCGGGGAGATCCCCAGGCTCTGTCAACGGTTGAGACAGCGCCCTGG      23103
23113      AGACCGTGAACAACAGATCCAGACCATCTGTGACATGCTGTGACAGGCCCTGGGATCA      23181
23173      ACCCCAAGGCGACCTGTCTTGACCGGATTTGCACTCAAGCAGAGAGTGGCTGCG      23211
23233      GGGACACCGGAGCGAAGTAGCTCCGGGCTCCA
...
```

```
23278  ACCACTTCTGTGTCGCCCTTGACCGGGCGCAGCCCGGGCGAGGACCAAGACCAGTCTGCTGTGAGC  |-----|
23338  CGGGCGAGCAGATCGGCTGAGCTGGCCGGTACGACTCAGTCGAATCTGCTGTGCCATGCGCG  |-----|
23398  AGCTGCAGGCTCGCCTGAGACGACAGGACCACACCGGGCATTCGCTGGCGGGGATGACTTGC  |-----|
23458  ACGAGTCAACACACGACGGCGGTGCGCCCGGCTG  |-----|
...
23607  GTCCGTAAGGACCCCGGCCGACGAGAAAGAGGTCAACCGACTACCAAGATGACGCTGTC  |-----|
23667  --|
23669  GTGGGACCAATCCCCGGGTGAGGCTGCAAGGCTCTTCCGGGGCGGCTCACAATCCGGTTG  |-----|
23727  TGGCTCGCGGTCAACGACAAGCGGCTACGATCGGCATCAGCAGCGGGGTCTATGTGAG  |-----|
23787  GTCCGCTGG  |-----|
...
23802  CTCGGCCAGTTGGGCTACCGGGCTAGAGGCTTGACCACTTGGTGGCTGCGGCTGCGGCTCAAC  |-----|
23862  TGCTCTGGGGAGCCGGGACCATCTTCACGTTGGTGGCTGGGACGCGTCAAGCGGCTCGAG  |-----|
23922  AAGGTCGAGTTCTACAAGAAGCGCTCTCGACTGGCCCTGTGTCCGACGACGCGGCTCCGG  |-----|
23982  TCGGCTATCGGGCTGGCTTCCGGGCTGGGGCTGGGAAAGCCAGGCCGTAAACCCCAAC  |-----|
24042  CTCGGCCAGGGCACCCCGAGTTTCGATCACC GGCTCAGCATCAACGACAACGACCCAGC  |-----|
24102  GGTCTCGGTGGGGCTCGGTTCCAGCTCGAAACCGGGCTCGTGGGCACTCTGCAGATCCCC  |-----|
24162  AACGGCGGACAGGTGCTACACAGCGGGCGGA  |-----|
...
24208  TCGGTGCCAGAGGGCTATCCGGGACGAGACCGTCCGCGAGTACTGCGCGGCGGACAGAGA  |-----|
24268  CGTGGCAGCCGCTGAATTAAGGTGCTGTAGGGCTCCACCTGGTTCGACAACACCTCGGATG  |-----|
...
24328  CCCAGAAGATGTGGGCTCCCGCGTGTGACGAACAAGAGATCTCCGGGGCTGACAMCA  |-----|
24388  CGCTCAACGACATCCAGTAGTGGCAGCACTGGGTCTCGGAGCGGTGAACGACGATGGGGTAA  |-----|
24448  TCGGGGCGGTTCGATACCGAAMCTGTGTACGATCTGGATCGGAGCTGAGGCCAGTTCC  |-----|
24508  AAGCTTCCCAACGAAGAAGCGCCACACACCTGTACTTTCAGAGACGGCTAATGGCGGGAT  |-----|
24568  CTCGTAGGGCTC  |-----|
...
24588  TTCCGGGCATCTGTGGAAACACCGTCCCATCTCTGAAGTCAAGCTTCGGCAGACCAAG  |-----|
24648  TGTGGCCAGGTTTCGACCCGGTCTGACTCCGCTCACGGCGGTCCGGCTACACCTACA  |-----|
24708  ACA  |-----|
...
24792  GGGGCCAAGCGGCTTCGGGGTTCTCGGGTGAAGGCCACGCTCAGGCGGTGCGCTGACA  |-----|
24852  TCCGTGGCGGTTCACACAGATCACTGGCGCTCATCGGCGCTGTGGTACCGCAGGCGCTG  |-----|
24912  GTTACATCTTTCGACAGACCGCGGCGGCGGCAAGGGGGCGACACCAACCACTTCT  |-----|
24972  CGGGCGGTGACAGCTCATGCTGCGGCTGTGCTGAGAGCAATTGAGGAACTGAGCT  |-----|
25032  TCGGAGGTAAGTCCCGGAATCTGCGGACATGGTGTACCGAGACCGGACCTACGACGCG  |-----|
25092  GCGCTAGGCGAGCTCACCCCAAGTGTATCGGGTACGACCACTGGTGGCGGCGCGC  |-----|
25183  CGGTGTCGCCGGTGGCCCGTGGCCAGGCGTGGTTCTGTGGCTACTAAGAGAGAGATGAG  |-----|
25243  TACGACTACGCGCTCGGGTACGAGGGCAACAGACACCCGATGGTC  |-----|
...
25309  CCGCAGCCAGCCTTCGAGAGGCTCGGGCGGGTACGAGGCGAGCCTTCCGACGATGATC  |-----|
25369  GACAACCCGACATCCGCAATCTGCAGATGCTTACAGGCCCAAGATCCAATGACCGTG  |-----|
25429  TGGACTGAGTGACAAGAAACCC  |-----|
```

```
...
25524 TTCCCTTCGTAATCTCTCTGACGGGCTGCTGAGGCCATCTGTACTTCAATCGCCATGC      |-----|
25584 GCGAGTGTGTGGCCCGGAGGGGCAACATCAGCTCTTGTGTGTGCGACCGGCTGAGCCG      |-----|
25644 CCAAGCTGGCTCCACGCGCCGAGGTGTGGAT      25622
...
25779 CTTCCGGCCCCCTTGTTCATCTTCTGTCGGGTACGCCATGTGCTCGCGGATCATCGCCGGA      |-----|
25839 CGTGAGGCGGACGAGTCAACGCGCCGCTTGGACTGACGCTTGTGGTGTGGCCGACGGA      |-----|
25899 TCTTCTGCGGACGCGGGGCGCCGCGCGGCGACGCGAGCTTATGCTCTCGCCGCTCAT      |-----|
25959 CCAAGATGCTCTTGTGCGGCGGATCTCGATCAAGCTCACCGAA      25943
...
26222 TGTTCACCAATCCCGGCCCAACGACGCCCGGACAAGGGCGGGGTCATCTCGGCGACCGGG      |-----|
26282 GTGTCCGCCCAACACCGGGGTAGATCCGCTTGCAGCGGTGCGTGTAGAGATCCTTGGTG      |-----|
26342 CCGCCAGGAGGTCTGCTGCGCTCGGGGATCCACTTCTTGTGTACTTCTCGACCGGTATGGCA      |-----|
26402 CTCGCCGACGCTTCT      26362
...
26438 GTCCACTCTCTGTTTCGATCAGCCGCTTCTCAGACGCGAGCCAGCGTTCCGGCTCCATC      |-----|
26498 CGGTGTGCGTAGTTCCTCGGCGCCGAGATACCGCTGCCCCGTCGATCGGGCTGACGTACGAC      |-----|
26558 GCTTGCACCTGACCGCTGCGCTGGGTCCGAGCGATCCCCATCTCTCCGCTTGTGCTGCC      |-----|
26618 ATGCGAAGAGGCTTACCGGAGATCGACCTTATTGCGACCTTCAGAGGCTTTCGCGTCC      |-----|
26678 GTGACACGAGGATTTTGTCCACTTCAAGTCCGACAGG      26661
...
26752 TGACACGCGGACCCCTTACTTCCAAACTAGCTACGCGGTTGATTTCCCGTGGCCGCT      |-----|
26802
```

```
26812 CCACAGGTCAAGGTGTTTTTATGCCGAGACCGGATCCCCGAGAGGGGGCGGAGCT
26872 TAGACGACACTAATGACCTGATTTATGTCTGCCGGATACGACATCTGCTACTT      |-----|
26932 TCGACCTCGACAGACGAAGAAAGCCCCCGGCTGCT      26913
...
27014 AGATTCTCTCACTGACAGGTGCGACTTAAAGTCCGAACGGCCGGAACCGTCCGCTGCG      |-----|
27074 GCGGTCTGCGCTTCAGCCTCAAGTTTCACCGCGGCTGAGGAGCTGTC      27064
...
27136 GCTCAGTGTGATGATGTCCTCGCTGTTATGACGAGGAGCATCTGTCGCGACATGGCG      |-----|
27196 ACCGTGGCCCTGAGCCGACACAGGTTGGTACGCTGAGGCGCTGTGATCTCTCGCGCTG      |-----|
27256 ATGTGCGTCCGCGGCAACGTGGCCACGCCACGCTTCACGSGCTATCGCGATGGTATC      27231
27316 GCGCGAATTCGCGCCGTTGTGGCTCCTCGCATGACGATCTGAGGTCTCTTACC      |-----|
27543 CAGAGCCGAGAGGTGCTACCGGGTATGCTCACTTACGATGGGGCGCATCGAGGCTTGA      |-----|
27603 CGGATCACTTCGACCGGATCAAGCTTGGGGCCGACCGCAGCGCTTGAGCGGCGCT      27593
27663 CGGCGATGTAGCGCGCAGCGCTTGTGTGCTCAGCGGAGAGTCCGGCGTCTGTGGA      |-----|
27723 TGAATGCTCTCTGCGGCAATTCAGCTCTCTTGGGGTAGACGACGCGCTCGATCCGGATG      |-----|
27783 CAGGATCAGGTTTCGACGAGCGGTAGCAGGGCTCTCGGGTACGTAGAGGTCGCTCGGA      |-----|
27843 TGAAGTCTTGGCGGTGCACTAGAGCAGTGCCTTGTGCTCAGCGTG      27784
...
27939 CCGCCCCAGAGAGTCCGCGAGGCGACGATGATCCTGACGTCCTCCGACGAGGCTCGGT      |-----|
27999 TGTAGCCAGTTCCGCGAATCTGTCGCTTGAAGAGACTGACCAACCTTCTCTCTT      |-----|
28059 CACAGTCGATGCTGCGGCGGCTGCCGTGGCATTCGCCGAGAGTACTGTCGCCAGTCCG      28030
28033
```

```
28068      28087      28119      28179      28239      28299      28359      28419      ...      28637      28697      28757      28817      28877      28937      28997      29057      29117      29177      29237      29297      ...
      GTCGACTCATCAGAAAGATGCGACAGTTGACGCCGCCCGCGGGGACAGATCAC      28149      |-----|
      GCCGTTCCGGGCCGGTGTAAACGGGGTGCAGAGCCGCCACCGTCACAGCGGTGAGGCCAG      28248      |-----|
      GCGGAAGGACGACGACGACGACGACTGCGATGTTCTTCATGAGGTCCTCTTGTCAGT      28245      |-----|
      TGTGCTTCCAGTTCCAGGATTTGGGCTTGAGCCCTTGCTTGTTCACAGACCGCTCGGC      28347      |-----|
      AGCTGGCCTTGTGCGATGTCGTTGGCCTGCTCCTTCGCGAGCGCCTGCTGAGGCTCG      28376      |-----|
      TGCAGTCGACGAGATCAGTCCGGGAGGCTCCGTTGAGAACCGCGACGA      28412      |-----|
      ...
      GAACTGTGGTCTCTCAGCCTCGTACATCGAACTCTTTCATGATTCGGTGCAGCGCTG      28687      |-----|
      GCTGACGCGCATCTCCAGGGCAACCGTCGCTTCAGCCACGCCCACTCGCGCTGCTGT      28777      |-----|
      TGAATCCCACTGGCTCTTGAAGCCACATCCTCGACAAGAAATGACTGTGACGCTGT      28777      |-----|
      GGATTCGGGTGTGTCTACCTGAGAGGTGATGCCCTCTCGGGGATGTACAGAGGACGT      28857      |-----|
      CCTGGCCGTCGAGAAGACGCCCGGCTGTGTACACAGCAATCAGGAAATGGTGATCG      28933      |-----|
      TTAGTGTACCCCTCTCTTGGTGAAGGCCCTCGAACACCTGGTCCACGACTGGTTGACGTG      28938      |-----|
      CCCCCGACGACGACGAGCCATGTCTCGAACGAGAGAGACGACTTGTGGGTCAAGACC      29036      |-----|
      TTGCGCTCGGAGTCCGACACGAAACTCGAACTCCAGTTGACCTTTACGTCGTGGCC      29107      |-----|
      ATTAGGCGCCGACCCAGGTACGGGATGATGTGGCCCGGAAGTGAAGTGAAGCCGGAG      29120      |-----|
      GCCGACCTTCGGGACATCGGCGGTGAGACGCCGATCTGCTCCCTGATGTCCTTA      29205      |-----|
      GCTGCTGAGCTTGTGATCGCCAGACCATCAGCTCTCGGTGAGCTTCCCGTTCGGTG      29254      |-----|
      CGGCGACGACACTAC      29254      |-----|
      ...
      29366      GTTCAACTGGCCACCTTCAGCTCGAAGTCCGGTCGAAGGTCATCCAGCGGACGAGC      29416      |-----|
      29426      CACCTCTTGGCTTCTTCCAGGTAGTTGGTGTCTGTCAAGTTGCCCTCCTCACTAATCAG      29484      |-----|
      29486      CGCCTTAGAGCGAGGCCCAAGAGGCTTTCACACTTCGGGGTGGTGGCATCAGCACCG      29507      |-----|
      29546      GACCCATCTCTCGGCCATCAGTACGAGCGATCTCGGAGCGGCCACAGTTGGCCTTCTCAG      29571      |-----|
      29606      CGGGACAGAGCGCGACGATCTCGTGTGATGGGACAGAGTAGGTACGGGGTGAATCCGG      29619      |-----|
      29666      CTTCTGAGAGCGCAATCAGGCGCTTTCAGAGTCAATCCGAGACGTGACTGATCTGGT      29683      |-----|
      29726      AGTTCAAGCGCGAGTACGTCCGAGAGCGGTCCACCGGACGCCGCCCATCGGGTTGA      29731      |-----|
      29786      T      29786      |-----|
      ...
      29929      CGGCCCTAGACCCGGGCCGAAGTTACACCGTCTTGGCATCTTCCGCTCCGGGTGTCTTG      29979      |-----|
      29989      GTGATGTGTGCCCGGAAGCGAGCGAGCTGTCTCATGAGTGAAGTCCGACCGCTCTTG      29979      |-----|
      30049      AAGCCCTGATCATCTGCTCCGGTCCGCCGAGAGCGGCCGACGAGAACTCTCTCGCC      30061      |-----|
      30109      TGGTAGTCGATCGAGG      30069      |-----|
      ...
      30134      GACCGGGGTTCAGCGAGGAAGACACGACGACCGTCCAGTCGAGCGGGCGAGGTCGG      30184      |-----|
      30194      CGGGATGCGCAGTATCGAATCGCGAGGATCCGGCGCTGCAACGGGTTGACGAACTGT      30184      |-----|
      30254      GGCACCGGTCTTCGAGTCCCGAGTGTCCAGGAACTCTTGAGACCCACAGAGTCCGCACT      30286      |-----|
      30314      TCCCGAGCTTCTTGGCTCTCTGACCATCTCGGCCAGGTGAGCTTCTGCTCCCGCAG      30372      |-----|
      30374      CGATCATGTGCCGGTAGAACCCCTTGTCCACCTTCCGCTTACCGGAGTGGTAGACGGG      30377      |-----|
      30434      TGAATTTGTGTCCACAGCTCTCGAAGCGCTCCGCGACATCTCGGTGAGCTTCTCT      30457      |-----|
      30494      CGATCCGTAATCTGTGAG      30457      |-----|
      ...
      30728      TCAGTACTCCGGGTGTACAGGTGATGATGTCGACCATCTTGGCCTTGTGCTGTGT      30728      |-----|
```

```
30778      GCTCCTGGCCAGCTTGGCCATGAGCCCCCTTGAGCTTCTCAAGCTGCTCTTCGAGATGA      30778
30848      ACTCCGCGATCAGCTCCTCCAGCGAGT      30819
...
30902      GGGGGTCCACCAGCTTCGCCAGATCTGCTCCAGAGATCTTAGCGCCAGCGACTCCA      30952
30962      TCTTGATCCGAGACCGCGGTGAGACACCTTGAGGTCGTACGAGGCTCTTGCAATCAGCA      31005
31022      TCTTGTTACGACGTCGAGCGCCTCATGACACCTCTCTCCGCCACGATCTTCAGATCCT      31039
31082      CGATGGCAGCAGCCAGGATTGGTCTGAGTACCGAACTGAGCTAGCGGCAGCGAAGT      31142
31142      CGGCGCTGTAGATGTCCAGGCCCGGTGGTTTGGGTGTCAGCGCGCAGGAGTTCTGATGAG      31143
31202      CCCGATGAAGTCCGCGAAGCCGTCAGATCTCTGGGTGTTCAAGCAGCTTGATGATGA      31245
31262      CGAGGTCCTTGGACCTCATGCGCAGCTGATCATGCTTC      31248
...
31316      TTCCGTTGACCGAGCCGACGAATGAACGGAGTGACTGTGCTCTGTGGACATTGGG      31366
31376      CCGACGACCTCTCTTCTTGGCCGTGCCGTGACGCGAAGCCGGGACGACGAGTGTGC      31405
31436      TGCAGGTCGTCAAGAGTCCGATCTCGCACCGAAGTTGATCTGGGGTTCGGTGGC      31442
31496      CCGGTACCGGACGGGTGTTCTTTCGACATGGGACCTTCCTTCATCAGGTCAATCGCC      31528
31556      ATGGGCGCGAAGGCCAACAGCCTGCACATCGATCGGTAGCGCTCATCAGCGGACGCGC      31592
31616      ATCACTACCTGATCCGATCGGCGCATCATGTGG      31616
...
31772      TCTGGCGGTTGCGCGGTTGCGGCGCAGCGAGCCAAAGCTGCTTTGACGAAAGCTTCAC      31822
31832      CGAAGTCCGCTTCGCGGTGTCCAGCTGACTCCGCAAGATGGGGTTCCTCCGACCGCAGGT      31872
31892      TCTGTTACAGCAGCAGCTGCGGGTTGACCGCAGTGA      31892
...
31942      GCGGCGTCAAGTCTTCTTCTTGATCTTCCTGCGCCAGTTCAACCTTGCGCCAGCCCG      31992
32002      AAGTCTCGGGGAGCTTGGTGTTCGCTCTCTCAGGTGCTTCTTC      32002
...
32185      CGAGATCGGCTCGCGTTGAGACACCTTGTAGCGGCGCTTGGCGAGTTGACGATGGTT      32235
32245      GCTGTTCCGACATGCTCAGCCCTCCGTAGCTGTAGGGTTCCGTGGGATGTCTCTGATAG      32245
32305      TGTGGAGCGATCTCCCGAGCTCCGAGACCAATTCCTGCGAGTTCCCGGATTTGCG      32312
32365      CAT      32365
...
32376      TCATGCCAGCGGTTCTTGATGACGTACCGCCAGCCCGGTGGTTGCCCTGACGACATC      32426
32436      GCGAGTTGTATGATGTCGCGAAGCGCCGAGCCGCTTCAGCGCTCTTCGCGGGG      32436
32496      AATCCGCGATGCTGAAGATGTGAGCAGAGTACTGTACGCTCTCTGGGCGCAAGACTGA      32504
32556      ACGTCCAGCAGACTCTCTGGCCTTGTGCGCGTCGAGCGCAGCAGCTCCGTAAGCGC      32560
32616      GGGGCAATGATGATCCAGCTCCGTGGGATGACGTACGCTGCGACACCAACCGAAGAG      32610
32676      CTCAGTGCCTGTGACGCTCCAGTTCCGTCAGAGCAGCGCGTGGCCTGATGATGAGAC      32689
32736      GTCCGCTGGAATCTCCAGCAGCAGCTCATAGGCCCAAGTCGAGATGTGTGAGGTAG      32738
32796      TCGACGTTCTCCCGGCTTGGGGTTCGGTGGTCAAGCTGAGGTAGCACTTCGGC      32798
...
32948      TCGCATGATTGACATTTCATCTCTCCGCTCAGAGTTGTGAGGGGCCAGTCGCTGTG      32998
33008      CTGGCCCCCGGTGGCTGTCAAGCGTGAATCAGCCGTTTCTTCTGTGG      33001
...
33322      AGGGCAACCACTTCTGGCCCTGGCCAGCTTGGCAACCTGTGTGGGATCAAGCCCAAG      33372
33382      TCGACGAGACCTCCGTGACGCTGGCTACTTGGGACACAGCAGCGCGGAGTACTG      33399
33442      CCGTCAACCTTGA      33442
```



```
...
33479 GCGCAGGCGCCACCTCCACGGTCTGGGTCTGCTGGSCAGGGGGAATCCACAGGCGACTCAG |-----|
33529
33539 TCGCGCTGCTGCTGCGCTTCTTCACGACGAGCGGTGTCAGCCGCGCGCTGGCGAA |-----|
33541
...
33624 TCTCATTAAATGGGGCATGCGCCGTTGGCGACTCTTCATCGACACCGCTCAGCGACGGCT |-----|
33674
33684 TTGATCCGACAGCCTCTGTACTGCTGCTGTTGTTGATTCGCTCGTAGG |-----|
...
33753 GGGAAATCGTGGAGCCCTTGATCAGCCCAAGAACTCTGCAAGATCGGGGGCGACATCC |-----|
33803
33813 ACGCCCTCGTAGGCGCTCCGGGTGACGTTGGCCGTGAAGCTCAGACGCTTCAGCGCTCGGCCAG |-----|
33824 33836 33857
33873 CACGCTGTGATGAGCGCCCTGGAAACGCAAGAGCTGTGTCAAGCGTCAAGCTGCTGTTGGCTGAC |-----|
33885 33894 33912
33933 TTCACGATGTCTCTGCGGCTCTCGGCCGACACCGTTCGGGCCACGGCCTCGACCAAGTGTCC |-----|
33947
33993 TTGCTCGGGA
...
34044 TTTCGACGTGTACCCGTCAGCCGCTACTGGAGCGGTGAGAACTGCTGCTTCTTC |-----|
34094
34104 AGCACCGAAGACCGGATGCGCCGATGAATACCGGGAAGATCG |-----|
...
34374 CCAACACCGAATGCGCCGTTGCGATCCAGAACTCTCGGCTCTTCGATCAGCGACGCT |-----|
34424
34434 GAGAACGTCGCCCGGATCAGGAACCGGCTCATCAGACGTTGGGCTCGGATCAGGTGATG |-----|
34437 34440
34494 TAGTCGCTCTTCCGCGCTCGTAAAGAACCGCGGAGTTGATGTGCCCTAGTTGCAG |-----|
34508 34521
34554 GGCCTCCACGCGCTCCAGCGTGAATCTCTCCACACGGGTTTGTGACAGACCTCGTTGGGC |-----|
34598
34614 TCACCGAGTTGACAGGCTGCTGTCCACATCCCGGCTGCGGTTGCGTACGCTGCC |-----|
34619
34674 T
...
34710 GAGCGCGGTTGAGCGGGTTCACAGAGGACCTTGGCCAGGGTCCAGAACGGATGCTCCACC |-----|
34760
34770 TCGACGAGATGTTGTTGTTCCAGTGAAGATCCGTCGTCCTTGAATGTTGTGACAGTC |-----|
34824
34830 TCGACCTGCCAGTCCGCCCAAGTCATCGCCATTCGAGCCGACCGGCAACACCGCT |-----|
34857
34890 GCCACCAACACTGCGGATGCG
...
35011 CCTCGTCAGCATCTTCCGACAGAGCACCGGCCAGAGCCGTTGCCGCCGACGCTTCA |-----|
35061
35071 GCTTGCACCCCGCCCGGACGCGGACACAGTCTGAAGCGCTCTCTTGTGGCGCAT |-----|
35123
35131 CATCCCGTATGTGGGTGTGATCAGTTCAGTTCAGCAAGCGGCTGCCAGC |-----|
...
35217 TCAGATGTTAGCGGCTTCGACAACTGACCCGCTCGGCCAAGTCAAGCGTATGTCGTGG |-----|
35267
35277 TCTCTGTGACAGACGATGTGACTTCCAGCTCTCTGCTTTCAGCTCGGGTATGACCGAGG |-----|
35288
35337 AATCGGTTGAGTAGTTGCGCCCGACGACCGGCTCCATGAGGCGGATGAAGTGAAT |-----|
35360
35397 TCGAAGTGTCCGAGGGCTCTCGATCCACCGCTCAACCCAGCATTTGAAACAGGTGCTGG |-----|
35456
35457 GCGTCTTGAAGCCCGACGCCACAGATGTCCACAGCGCGGAGAGATCTGAATCCCGC |-----|
35461 35470 35477
35517 ATGATGCGAG
...
35693 TTACACCTGTGTCACTACTTCTCTCAGTTCATGCTTGAATAGCTTGAATAGCGTCAAGCC |-----|
35743
35753 GAATCGCCCTTGGAGAACGAGACACCGTACTCATGCGCACCGAGTACAGCCGGTGAACAG |-----|
35755 35785 35809 35812
35813 ACGCCCTTCTTGGCCCGGAACCGGCTATGCTGTGCGGAGGTACTGTGCTGTGCTG |-----|
35814 35817
35873 CGTGGACGATGACCTTGTGTCCTTCTTCAAGTGGGCTTCCGCTGCGGCTAGCCAGGGG |-----|
35878
```

```
35933   TGAACAGCGCCGCCAGCTACTGCTCCAGGTGCTCCTCGGACCACTTCTCCAGAGCATCG      |-----|
35993   GCTTCTGATGTGGGAACAGCTCCG                                     35961
...
36209   ACCTTTCGAGATGTCAGGGAGATGGCTGCTTGTGCTCCAGGGCCGCTCTTGACGATAGCC      |-----|
36269   TTGCGGTTGTCAATCTCGCTCAACACCAGGGGTGATCGCGATGATGTGACGTGCTCGGTG      36259
36329   |-----| |-----| |-----| |-----| |-----|
36329   AGCGACTTTCAGGGCTCGCTTCAGCATCGCTCGGACGCCAACCCTCTCTCCGGGACGAG      36382
36329   36329 36353 36382 36385
- |-----| |-----| |-----| |-----|
36389   CCGTCGTGTACCGAACCCTGATCGCTCTGCGTGCGCTGCTTCTTGAGCCAGTTCC      36391
36390 36410 36423
36449   TTCAATCCGACGCGCAGAGATGTCAGCAGGATGCGATTGTCAGTCACCCCGACGCG      |-----|
36452
36509   TCCTTGAGAGTTGTCGACGAGTAGTGGCTGCGCTCTGGAAGATGCCCTTGCGCC      |-----|
36514
36569   G                                     36514
...
36638   AGCTTCTTCTGAGTAGTACGACTTTTCAGAGTACAGACCACAGCTCTCTGACAGATTCG      |-----|
36688
36698   TCCGCTGTCAAGTTGGTCTCCAGCCGATCATGCTTTGGCGGGTGGCCCGCATGAAGAC      |-----|
36728
36758   TTGTTGATGTGCGCTGTCAAGGATCACTTCGCTAGTACTCTCCCTGTGACAGAGAC      |-----|
36766 36812
36818   GGTGAGAGTTGCGGGGATCTCTCGGCGAAGACCTCAGGAGCTCGCTG
...
36922   GTCAAGTTTCAGACCTCCAGACGTGGCCGTCGACCGAAGATCGGCCCTCCAGCATGGA      |-----|
36972
36982   ACCAGCTCGGGCTTGACGTGGTTGCCGTCAACCGCTCAACAGAACCA
...
37195   AACCCTGCATCCGCTTGTACGCGGTGTACGCCGACTTTCGCGTCACCCGGACGCCGCA      |-----|
37245
37255   CGGTGCGCGGTGCGTCAAGATCCAGCCCGGAGCCACTTTGTAAGACT
...
37384   GCGAACTTGTGACAGATAGTCGAACGCGCGGCTGTGTTGCCCTTCGTGAGCGCCAGCA      |-----|
37434

37444   GGGCGCTGTACACGGCCCGCAGCGGGGCCAGGAACCTTCGCTTGG
...
37525   TTCTCCGACGAACCTTGGTCCACCGAGACGGGCTTCGGGTAGTCCATCAGGTCCGATG      |-----|
37585   TGGACGACCTCGTCGGGCTGGGTGTCGCGGATGAACCCGACACAGCCTTGAGTGTG      37575
37585   37587 37608
37645   CGGTATCGAAGCGGATCTGCGTGTGAGAGATGACGACGATGCGCTTACTACCCAGTAC      |-----|
37677 37680
37705   CTCACGAACGGACCCCTCGAGATCTTGGCGTGTGAGCTTGAAACCGTCGCGGTGCC
37765   CCACAGGGGCTGTGTGTAGTGGGGTACCGGACCAAGACGGCTTGCCCGGCTCCGAC      |-----|
37770
37825   G
...
37922   CCGGCCATGTGCTGTAACGAATCACGCTGTGTAACCGGTTGACCTTGACCTTGCGATTC      |-----|
37972
37982   ATCAGATCATCAGGTTGGCCACGTCGATGTCGAGATCTCTTCTCCAGGTAGCCGAG
38042   AACACAGGGGAGATGCTCTTGAAGTTCTCTCGCGGGTCCGTAAGTTCTTGTGGCTCA      |-----|
38050
38102   CCGTGAATGAGCGGCTGAGCTTCTCGAGATGATTCACACAGTCCGCTCCGATG      |-----|
38112
38162   CGGTGT
...
38201   GATAGCCGATATCGGCTCGAATGTCGTCAGATGACCGGAACCGAGAGCCTTCA      |-----|
38251
38261   GTACGTCGTGACGTACGTTTGGCTTCGGGCTTGGCGGTGACGTCGACGTCGTAAT      |-----|
38289
38321   CGAACCAGCGTCTCCAGGAACCTGTGATCCGCTGACAGGCGGCAACCCGCGTGG      |-----|
38328 38331
38381   TGTAGATCGTANCGGTGTGTAACATTCGTCGATCAGATCCTTTCAGACAGCGCTTT      |-----|
38404 38432
38441   GCCCTGCGATGTGACTAGTAGTGTGACATCTCGCTTCTTGCGCATGGGATGATTCGTC      |-----|
38501   GTTCGCGAGCGCTTGTGGCCACCGACTTCGGAACTCCATACCGGGCTGTGCGCTGACG      |-----|
38508 38549
38561   CAGGATGTTACGTTGCGATACCGAGGAACAGCTCGCGAAGTGGGCTTCCACATCTG
```

```
38621 AGCCCTGGAAAGCCGACCGTCGGGATCCCGCACAGCTCAGCGTAAATCCGTGATCTC 40211
      38630 |-----|
38681 ACCCTGGGTGATGCGCATGTCTCCGCGAGTAGCGGGTCAATGGACCGCTGTTGACAGCCG 40217
      38691 |-----|
38741 GGGCTTATCCCGCGCATGATGTCATGTACTTCGGCTTCCGTCGTCCAGAGCGGATACCG 40260
      38780 |-----| 38791 |-----|
38801 GATCGCTGCAACGACCGAGTTACGCCAGGGCAGACCGCATGTACGGGATCGCCAGGCA 40266
      38802 |-----| 38829 |-----| 38850 |-----|
38861 GCCTCGGTACATCTCATGACGAGGAGTGGATCTCCACGAAATCC
      |-----|
...
39244 GCGATGCCGGTGGCTGTCAAGTTTGACTCCTCAGCTTGTGGCCTCGATGGGGCTATGC 39294
      |-----|
39304 GCTCCCGGATGACCTGAGCGCGCGGGGCTTACGCAAGTAGTCGATCGCTCGTGAGCA 39348
      |-----|
39364 TCTCGATGCAATCCCGCGCCGACCGAGAGTGTACTTGTTCACATGTCGACAGAGCACC
      |-----|
39424 CTCGAACGATGCCCGTCTTGTGTGATGTCGTACGATAGCCGTTTCTCTTTCCGTTGC 39428
      |-----|
...
39597 GCAGCCTTGAACCTCCCGGTGGTGGTGGCGCACCGTGGCCGGGATACTTGGCGTCCCGT 39647
      |-----|
39657 GTAGCGGCAACCCCTTGGGCCACACAGTCTTTGACGGCTTCCGTT
      |-----|
...
39706 GCTTAGATTTCTGACCCCGGTACCCCGAGTCCCTTCGCGGCTTCCCATGCTCCCG 39756
      |-----|
39766 TTCGGTTGTTCATCAGCAGGAGAAGTACGACAGCACCGCACAG
      |-----|
...
39818 TGTGTGGTGGAGCGGCAACAGCACCGGTTTCATCCGCACTGTGATGATCGCCGTCGG 39868
      |-----|
39878 GAACAGATCCAGCAGAGACCGAGCGGATGAGATCAGCCCGCAGCTCCGGGTACGACAGAT
      |-----|
39938 CCATAGTGGAAAGCCGTGACCTCCGTAGAAAGAGCGGTTGTGATCTTGGCCGTGTCAG 39971
      |-----|
39998 GCCGGCCTCGTACGGAAAGATCTGATCATGACGCCGTGGAGCAGCGCGGTGGTACGCC 40013
      |-----|
40058 GTGGCGGTACA
      |-----|
...
40161 AGAAGCTCAGACTTCTGCCGGTATCTCGCAGGTGCAWGGCCTGTGACCGTCCGCCAGC 40112
      |-----|
      |-----|
      |-----|
```

```
41125
...
41190 CGGGTAGCGGACTCCGCTACTTCACTTGAATCCACTCTTGGCCAGTCGGGTCGACGTTTC
|-----|
41250 TCCGAGACGTTGGGGCCAGGCCGGTGAATCCGCGCCGTGACCTTGTGTCATC
|-----|
41310 CGAGGCTTCTCGAATCCGGGCTCTCCGGGTGATCGCTTGTGCTCCAGACGTCGGA
|-----|
41370 CGCGTCTTGAATACGCGGCCAGCACACTGTGTGTCACAGGGGTTGGCCTTGGGGGCATG
|-----|
41430 GCGTTCGAGTGTGATCTGGGGTCTCTTGGGTGCTGCTCAAGGATGATGAAAGG
|-----|
41490 GTTGGGGACATGAACCGTTTGTGCTTTAAGCCAGGGCGGTTCCCGTCCGCTTACC
|-----|
41550 GAGCAGACGCGAGAGCCTAGGTACGTACGTACGCGCCCTCGCCCAACCCCTTGGT
|-----|
41610 CAACGAAGCTCGATGTCGGGAGATGACTGCGGCTTGAAGTTAGCCTGTGATAGACG
|-----|
41670 TCGGAGACGTTCCGCGCTTGGATCTGCTCCAGAAATACGTGATGTGCGACAGGCC
|-----|
41730 AGGAAGTCTTCTTGAACGTGGCCGCTTCTGCTTGACG
|-----|
...
41871 GGGATCTCGAAGTTGTACGGCCCTTCTCAAGTTCTCTAGGCCACGTCGGCTGAG
|-----|
41931 GTACACCCGACCAAGCCAGACCGGGCGAGGGCGGAGTGGCTGGGATGGCTTTC
|-----|
41991 TTGATGTGCGCTACTTTCGTGGTGGTGGCTGTCAAGTCAAGTCAAGTCAAGTCAAGT
|-----|
42051 CATGGTGTCTCCGATGAATCCAAAGGGGGGTAGTCAATCCCGACGGATCAAGCTTGGC
|-----|
42111 ACCCGGTTCTTACCGTGGAGACGTTGAGCGAGTCCGGGCGAAGCGCTCCGACTCGG
|-----|
42171 ATGGAGCGTAGGACCATCTCGGGCACAGCCCGATCTGACCTTGATCCCTCTTAGCGG
|-----|
42231 GATCGCTTGTGTCGCGCTGCTTGTACGGGCGGTGACGTGTGAGGCCCGGATCAAGCAGA
|-----|
42232 42240 42246
42291 GCCAGTCTCCCTGGCCATTTCTGTCAGGTAGTCAATCAGGACCTCCAGCCGCTGAACGG
|-----|

42351 GTGATCCCCGTCGCCGACTCGGTGCGGACGTTGGTGAATGTTGTCCAGACGATCAGCC
|-----|
42411 AGGAAGTCTCTGTAGACCGCGCTGTACGCTCGAGGCGGCTTCTGATCTGCTCCAGCA
|-----|
42471 CGAGATGCTTGTAGTTGAACCGGATGGGATCTCATCATGAGTGTGGCATGACTC
|-----|
42531 CTCGATGCTCTGTTCGCCGACTGCCGCGCTGATGCTGAGCGGCTTCCATCCGCTCAGAT
|-----|
42591 GGAGACTGACCGGAGATCTGGGTGAACGCGTGGAGTCCGCCGAGAACTACAGCGTG
|-----|
42651 TACCTTGACTTCAAGCGGTAAAGCCAGACAAAGCCGATTTACAGT
|-----|
...
42782 GTGGGTACCCCGCTTCCCTCGATGTACAGGCTCTGCTCGGTGTGATACGCTTCC
|-----|
42842 TCTTTTCTCTGTGAGCATCTCCGGAGCATCCAGCAACCGGGGATGCTGTTGAACCTC
|-----|
42902 ACAGACCTCTCTGATGTCGTTGACATCTGCGGACCTGCTCATGGGACACAGTGT
|-----|
42962 GCCAGGCTCCAGCCGACGATGATCTGCGGTGTCATGTGTCAGCCGAATGATTC
|-----|
...
43032 GTGGCGGTGAGTAGATCAATCTTATCCCGAGCCGCACTGCTGAAGCTTCTCTC
|-----|
43092 GCGGTGTGTCGCGCGACGTAAGGGAAGTGGCTCAGCAGCAGTTGTTGACCGCTCC
|-----|
43152 GTCGA
|-----|
...
43177 GCCGTGACATGTAATGACACAGTTTCCAGTACAGCCGGGAGACCGGGGAGGTCGCG
|-----|
43237 TACATGGGTGAGGCCGATATGTTGGCCAGATCAGCCGCTTGC
|-----|
...
43438 CGGTCTCCGCTTACCTTGGCGGTGGCGATGTAGTCCGAGGTGAACGACGATCTC
|-----|
43498 ACGGTGGCTCTTACGCTCGGCCGACCGGTGGCTTCCAGATGAA
|-----|
...
43561 CCGGATGAAGAGGTATCCCGCTGATCTTGTGCGGAGCCCAACAGATCGCCAGGA
|-----|
```



```
...
46019 GCGAATTGGCTTTGGGGTTTCATGCATCCAGTATCCAGTAGTGTCTTTGACGCTGAAG      |-----|
46079 CGTGGGTCGCCAGTCACAGACTTTCGATGGCCCGGAGAGATCTCGCGTCTCCGCTTGGCGGC 46069 46077
46139 TCCATCCCTTCGCTCTGACGCTTATCTGATTCCTTCGCGCTTGGCAGGCGGTGCT      |-----|
46199 TCCAGGGGTTTCATGCCAGCGCGGAGCGGCGCGCATGGCAACCATGATCTCGACGTTCA      |-----|
46259 TGAGGCCGGAACCTCCACAGTACCAAGGGCTTCCGCCGGAGTATGATGTACCTCCATC      |-----|
46319 GCCGAGCTGGTGCATGACCAAGCGGCTTGGCGGAAGCGGCAGCTTCGAGGTGGTGAATGAG 46321
46379 GATCTGGCGATGGCTGCGCCGATACCTTCTGATTTCTTGGCGTCCGCGTCCACGGTGGCAGT      |-----|
46439 C      46384
...
46463 TCTCAGTAGTCGAGTGTCTGGAGTTTGAAGCCTTCGAGGTACCGACCTCTCGGCTTCA      |-----|
46523 GCGGAGCGGTTGCCGCGCCAGTCTGTCGGGAAGCCTTGGCGCGTTCGCTTACAGAGCA      |-----|
46583 CCGCAGTTGGGACAGTCCACGTCGCTCTGGGCTCGGTAACCTCTGCGCTCATGGCCGAC      |-----|
46643 CGTGAGCAGTCCACGCGCGCTAATCGCCGTCATGATGAACCTTGGCGCTTCAAGCACA      |-----|
46703 GCTTCAGCAGTAGCGGTAGTTGTTGGCCATGATCAGTCCCAAGCTCTGTCGATGGCTT      |-----|
46763 GGTGGGAGCCTCAGGSAAGGTACGCCGTGCTGACGCGTCTGAGTAGSTCCAGTCCG      |-----|
46823 CGATGTCGTCGCGCGACCCCGAAGCTCTTTCATGATGAGGCTCCAGGTTGGCCATCC      |-----|
46883 ACAGGTGAAGCTCACAGTACCTCATTCTTGGTGGTGAAGACAGAACGCTCTTTCGT      |-----|
46943 CCGTTGATGCACAGACCGCACTTAGCGGACCGCATCCCTTGGCTGATCAGGTGGAAG      |-----|
47003 TCTCGGTTGTTCAAGTCCGGGCAAGCGCAGGCGGTGCGGAACGTGTGCTTGCCTTGGCGG      |-----|

47063 AACGTCGTGTGACGTAGGCCATGTTGATGCCCTTGCCTTCCAGSAAACCGGCGCACGTC      |-----|
47123 ACGTTGTCGGGATCCGCGCTGAATACAGCCACAGAGTTGCAGACGCTTGTGCGTCAAG      |-----|
47183 TAGACCGCTGGCGTCTGAACCCCTGGTGAAGCCCGAGAACTGCACGCTCGTGAAGTCTCGG      |-----|
47243 ATGACACCGCGCCACGCTTCGACAGTAGGTGGGCTGAAGAACTGCCATCCAGTGAATA      |-----|
47303 CGGAACAGCTTCGAGAGCTTCTTCTTGTTCGAATCCT      |-----|
47342 ACGAAGTCTGCCATCATCTCGAGACAGAGGTCACGGTGTCTGTCAGTACAGCTCACGC      |-----|
47402 AACAGCTCCAGTGTGTCAGCAGACGACACTGCTCAGCGCTTTCGGA      |-----|
47507 GAGAACCTTGGCCCACTGGCGAGAACCATGCTGTTCCGGAATGCCATACGGTGGCGTGGCG      |-----|
47567 CCTTTGGTGAAGTGAAGCTAGTGTGATCTTGGCGGTTCGATCGCTTACGATCCC      |-----|
47627 ATGTCTAGAACCTTCCGCTGCTGCTGCAAGTCAAGCGGAGCGCAATGAT      |-----|
47722 CTCGGAGSACATAGTGAATCCTTCTCTGGTGGCTGTCAAGTCTCAGGCGCCAGCAGCTT      |-----|
47782 CGTCGCGCGGGGGTGGCGAACCTTCGAGCGCTTCTTACGCGGTGCTTGGCGATGACGT      |-----|
47842 GTCCCAAGTGAATGATGATGTCCTTCGACAGGAGGCAAGCTCAGCTCTCAGAGGATCTCGA      |-----|
47902 GGTGGGACGTTGACCACTTCGCGCTTGTCTGTAAGTCAATCCACCCCTTGTGCGCTT      |-----|
47962 CTCACGCTCAGCGTCGACAGGTGCGAACCGTGCAGACGCTCAGGTGCGACGAGAGGTGCTT      |-----|
48022 GATGG      |-----|
48150 AGTCGTATGCCGCGATACCAAGCTCAGTATCCGCCAGACCATATAGCTCAGCGATATT      |-----|
48210 CCATGCGCGCAGAGCTACCGGTAATCTTTTGTCTTCCCGTGGGT      |-----|
```

```

48431 CGGTGGCTGTCAAGCGCGCGTTTGGTATGCGATGATGCGCTGTTTAGCGCGCTG 48481
48491 GCATTAAGGCGCTACCCGCTGCTGACTACCGGCTCCAGTTGGTATGACATTCACTTAGCG 48497
48551 TATCCGCTCGGTGGCTGTCAACCGGAGATACACACCGGATTTTCACGGGATCCGCGCGCAT 48609
48611 CGTTTCCGATCCGCTATGCGCGCTGCTGTGCTTGCTGCATGACACAGTAAACACC 48614
...
48818 ATGTGCCCCGATATGGGTGCTGTGAGCTGCACATCCGCGCGCTGCGGGCTGTGCTCAGCG 48868
48878 TGTTCGCTGTGCTGCGCGGGTGGCTGTCCAGTCCAGTCCGCCACGAAATAGCGAATAAGCACC 48879
48885 48888 48907 48885
48938 GTGCTGCAAGTGTCAATAGGCTGCTCACTATGCGATGCGATGCGATGCCCTTGACACAGTG 48938
...
49058 ATGGGACAGTGTGCTGTGAGTTAGCTGTAGCTGAGCGCGCGCATTCGCATGTCGCAG 49108
49118 GTGACGACGGTATGGCTGATGTGTCTGCGTTGCTGGGCAATGATGTCGCCCTCGAGGC 49118
49178 ACGGCTGCGCGTGGCGTTTGTGCTGTGCATGCGCATGTCGCGACGTCAGCGGGGTAGGGGGG 49190
49238 TTCCCCCGACGGCGCTTCTCTGACCGGTGCTTA 49249
-----
40 matches found in sequence:
aav44440 : Mycobacterium tuberculosis antigen Tbh-4-Xp1 DNA.
(from "mycobacterieng.seq")
TOIG of: aav44440 Check: 3420 from: 1 to: 2138
ID AAV44440 standard; DNA: 2138 BP.
XX
AC AAV44440:
XX
DT 09-NOV-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen Tbh-4-Xp1 DNA.
XX
KM Tuberculosis; infection; diagnosis; antigen; Tbh-4; ds.
XX
OS Mycobacterium tuberculosis strain H37Rv.
XX
OS Mycobacterium tuberculosis strain Erdman.
XX
FH Key Location/Qualifiers
FT CDS 204..1586
FT /*tag= a
XX
PN W09816645-A2.

```

```

PD 23-APR-1998.
XX
PF 07-OCT-1997; 97MO-US18214.
XX
PR 13-MAR-1997; 97US-0818111.
XX
PR 11-OCT-1996; 96US-0729622.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX
DR WPI: 1998-251292/22.
DR P-PSDB: AAM64367-68.
XX
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX
PS Claim 4; Page 188-189; 250pp; English.
XX
CC This is a composite sequence of overlapping DNA clones Tbh4 (see
CC AAV44381) and Xp1. Tbh4 was isolated from a Mycobacterium
CC tuberculosis strain H37Rv expression library and Xp1 from a M.
CC tuberculosis Erdman expression library. The composite sequence
CC codes for a 460-amino acid polypeptide (see AAM64367). The reverse
CC complement also contains an open reading frame (see AAM64368).
CC Recombinant Xp1 protein stimulates cell proliferation and
CC interferon-gamma production in T cells isolated from M.
CC tuberculosis-immune donors. The invention relates to methods for
CC diagnosing tuberculosis. It provides polypeptides (see
CC AAM64291-W64379) comprising antigenic or immunogenic portions of M.
CC tuberculosis antigens, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers.
XX
SQ Sequence 2138 BP; 437 A; 624 C; 738 G; 339 T; 0 other;
AAV44440 Length: 2138 March 5, 2002 14:19 Type: N Check: 3420 ..
Found using 'seq2-3' (Papua403.key)
...
32 CAGCTGCCGATTCGCGGGTTTCCCGACCGAGAAAGCGCTACCGATGCGCTGCCG 82
92 AAGTAGGGCGATCGTTCCGGATGCGGCGCATGACGGCGCGCATCAAAATTAGTCAGGAA 87
98 103
152 CCTTTAGTTTACGACGAGATATGGCTATAGCACTAAGGAGATGATCCGATATGACGCA 163
166 171
212 GTCCGACAGCC
...
283 ACCCAGCACTGATGTGCCCATCACAACGTCGCACTCAACGCGCGCTAAACGCGCGCC 333
AACAGCTGATATGTCCCGCGACACATGCGGGAATACCTGCGCGCGCTGCCAAGAGC 338
403 GGCAGGCTGTGGGACCTCGCTGCGCAACGCGCGCAAGGCGTATGCGAGGTTGATGAGG 406
411 418

```

```

463      AGGCTGCAGCCCGCTGGACAGCAGCGCGAAGAACTGTGACGACAGATTCGCCCGGGG
      |-----|
      483 488
523      CCGTCGAGGGGACAG
      |-----|
      488
...
580      GTGAACCCAACTTCAATGATCTCAAAAGAGCGCAAGAGCTCGAAGCGCGACCAAG
      |-----|
      630 635
640      GCGCATCCCTCGCCGACTTTGCGGATGGGTGGAACACTTTCACCTTCAGCGTGCAGGCG
      |-----|
      644 649
      686 691 696
      699
700      ACCTCAAGCGGTTCCGGGGGTTTGACAACTGGGAAGATGCGGCTACCGCTTGCGAGG
      |-----|
      701
      704
760      CTTCGCTCGATCAACAACGGCAATGATGATCTTCACATGCGCAATTGAGCGCTGGATGG
      |-----|
      807 812
820      CCAAGCAGGCTCAATATATGTCGCGAGCTGCACGTGTGGGCTAG
      |-----|
      807 812
...
1128      GTGACTCCCGGTACCGGGAGTCCACGCGACCGATGTTCCGCTACCGGATCCCGGGT
      |-----|
      1178 1183
1188      GGTGGCCTCCCGGCTGACACGCGCGCCAGCTGACGTGGGCTGGGCGGGAACCGCGAGCG
      |-----|
      1220 1225
1248      -| |-----|
      |-----|
      1249 1254 1259
      1277 1282
1308      CCGTCGGCGCGTGTGGGATCCGCGATCGGGGCGCGCAATGCTGCGGCGCGCTGGCGCT
      |-----|
      1313 1318
      1338 1343
      1362 1367
1368      GGTGACATTGCCGGCTTAGGCCAAGGAGCGCGCGCGCGCTGGGCGCGCGT
      |-----|
      1407 1412
1428      GGCATGGGAATGCCGATGGGTGCCCGCATCAGGGACAGAGGCGCCAGTCCAAAGGT
      |-----|
      1470 1475
1488      TCTCAGCAGAGAACGAGCGCGCTTACACCGAGATCGGCGATGAGACCGCGCTCATTT
      |-----|
      1505 1510
1548      GGTACCGCTGGCGCCAGACAGTAAAGAGTGAAGTGAAGTGAAGCAATTGGACCGCGC
      |-----|
      1558 1563
1608      ATGTCCCGCGGGGTGACGCTGCGCGCGCGGTTTCACTAGTGGCCCTAGACGAGGAGCTCA
      |-----|
      1610 1615
      1623
      1624 1629
      1660 1665
1668      ATCAGATGAACAGGATCTCTCCGGGCCACCGACGAGCGGAGACCGCTGGAAGTGAACGA
      |-----|

```

```

1728      TCAATGGCACCAGTGTGCTCACCGCGCTGCGCATCGAAGATGTTTGTGAAGAAGCTGG
      |-----|
      1728
1788      GTGCCGAGCGGTGCTCAGCGGGTCAACGAGCGCTCCACAAATGCGGACGCGCGGCT
      |-----|
      1819 1824
      1843
1848      CCGGCTATACGACGCGCGCGCGCGGCGAGCAGTACCGCTGCTTATCGGCATGTCCCGG
      |-----|
      1848 1856 1861
1908      CGATGACGAAGAAATGCGCTAAGCCCATTTGTTGGCGGTGTAGCGCATACGACCGAATG
      |-----|
      1949 1954
1968      AGCGCCGCAATGCGGTCAATTCAGCGCGCGCCGACACAGCGCGTGAATGTCATGTT
      |-----|
      1968 1973
2028      TTGACATGATCGCGCGGTTGCGAGGCGCGCATATGCTCTGTCGCCAATATTGCCGAG
      |-----|
      2054 2059
      2069 2074
2088      CTAGCTGGTCTTAGGTTGCTGTTAGCGTTAGCTGTTAATTATGACGTCGTTACCA
      |-----|
      2125 2130

30 matches found in sequence:
aav44442; Mycobacterium tuberculosis antigen RDIF5 DNA.
(from "Mycobacterium.seq")
TOIG of: aav44442 check: 2013 from: 1 to: 1923

ID      AAV44442 standard; DNA; 1923 BP.
XX
XX      AAV44442;
XX
XX      09-NOV-1998 (first entry)
XX
XX      Mycobacterium tuberculosis antigen RDIF5 DNA.
XX
XX      Tuberculosis; infection; diagnosis; antigen; RDIF5; ss.
XX
XX      OS      Mycobacterium tuberculosis strain Erdman.
XX
XX      FH      Key
XX      FT      CDS
XX      FT      Location/Qualifiers
XX      PN      WO9816645-A2.
XX
XX      PD      23-APR-1998.
XX
XX      PF      07-OCT-1997; 97WO-US18214.
XX
XX      PR      13-MAR-1997; 97US-0818111.
XX      PR      11-OCT-1996; 96US-0729622.
XX
XX      PA      (CORI-) CORIXA CORP.
XX
XX      PI      Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX      PI      Reed SG, Skelky YAM, Twardzik DR, Vedvick TS;
XX      DR      WPI; 1998-251292/22.
XX      DR      P-PSDB; AAW64373.
XX      PT      New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX      PT      to develop products for the detection of M. tuberculosis infection
XX      PT      and diagnosis of tuberculosis
XX

```


PS Claim 11a; Page 197-198; 250pp; English.

XX This DNA sequence codes for an antigenic portion (see AAM64373) of
CC Mycobacterium tuberculosis antigen RDIF5. It was isolated from a
CC M. tuberculosis strain Erdman genomic DNA expression library using
CC rabbit anti-sera raised against M. tuberculosis fractionated
CC proteins. Phage plaques expressing immunoreactive antigens were
CC purified. Recombinant RDIF5 was shown to stimulate cell
CC proliferation and interferon-gamma production in T cells isolated
CC from M. tuberculosis-immune donors. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAM4291-W64379) comprising antigenic or immunogenic
CC portions of M. tuberculosis antigens, as well as DNA sequences
CC encoding such polypeptides, recombinant expression vectors and
CC transformed or transfected host cells. Also claimed are methods
CC and diagnostic kits for detecting M. tuberculosis infection in a
CC patient using these polypeptides, antibodies or oligonucleotide
CC probes and primers.

XX Sequence 1923 BP; 344 A; 595 C; 643 G; 341 T; 0 other;

AAV44442 Length: 1923 March 5, 2002 14:19 Type: N Check: 2013 ..
Found using seq2-3' (peppu403.key)

```
1      TCACCCCGAGAGAGTCGTCGTCGACGACATCGACTCGCTCGATGATGATCGAGA
      14 19 24 29
      61      TCACCCCGAGAGAGAGACGACAGTACGCGCTCAAGATCCCGACGAGACCTCGCCGCTC
      65      87 92
121     TCGCTACCTCGGTGAGCTGTGTCCTTACATCCAGAACTCGAGAGAAACCCGAGG
      135 140 141 146
181     CGGCTCAGGCGTTCGCGCGGCAAGATTGATGTCGAGAACCCCGATCGGCACGACGATC
      188 193
241     GGTGCGTTTACCCACATCGCAAGCTCGAGCGCGCTGCTTCTTGACGCTGACCA
      270 275 282
301     GGTGCGGTTCGCGCGCTTCCAGCAAGTGTCCACACACAGAAAGGACCTCGCGAAA
      310 315
361     GGTGA
421     GAGTTGCGGGGGGCGAATTGCGGCAATTGCGTCAAGAGCCGAGCGGATCCCGGCGCCGCC
      471 476
481     CGGCGTGGCTGTGTGTTTTGGGCGCCCGGATGGCCACGACGAGCAAGATGGGCGGAT
      523 528 535 540
      526 531
541     GAACAGCGCCACGCAATCAACAGACAGAGATTTCACAGCAAAACCTCTGTACCGCTG
      545 550
601     CGCCGCGGTTGTGATCGGTCCGATATCATGATGCGCGCTTAACGTAACAGCTTTGCGC
      633 638
661     GGACCGGGGGTCAACAACGGCGAGTTGT
```

```
...
775     AACTCAACGGCAACGGTTTCTCAGTCCACCAGCTCAACCTCGACCCGCAATGCTGTGTA
      825 830
835     CGTTTCAGCCGCGGCGAGTCCGAGTCCAGAGCTTTGGCCCGCA
...
1058     AGCGGCTCAAGACCAATAGTTTCCACAGCCGCGTGCAGATCAGATGAGACGCGCAGG
      1108 1113
1118     TTGAGCGCGCGCACGGCGGCTGTGCTTCCGTGTCAGGTCCGA
...
1166     CCGGCACACGACACGCTGTGTCTGTGTCGATCACCGCGCTGTGCGATCGACGCTTCCC
      1216 1221
1226     GAACGATTTGCTGCTGTCACAGCGGGGCAAGGACGCTTGTGCGCCGCGCAGCAACCGAGC
      1227 1232      1256 1261
1286     CTTCCGAGACGAGTTTCACACCGGTC
1312     GGGCGCGCTCAATCTCGATGCGCCCATCGCGCTGCGTGATTCACCTGATGCTGCTCCG
      1362 1367
1372     CGCAAGCCAAAGGCGCTGCGCAATCCGCTTGGGAATCACACGACGCTCTGCGACATCGATG
      1382 1387      1412 1417
1432     GTTGTTCGCAWGTAGGAAATTTACATCGCACGCTTCATAGGCGTGTCTGCGCGGGAT
1492     GTCGGGAGATCCGCTGACCTATCGAAGCATTTGTTTCGGAATGCTGAGGAGACGTGCG
      1497 1502      1517 1522
1552     GTGCGGCTGATGGGTGTGATCCCGGCTTGACCCGATGCGGGCTGTGCTCATCGAGAT
      1596 1601
1612     GGGCGTGTGCGCAGCTACACCGCGCTGATGTGACGTGTGCGCACACCGCTCGGATGCG
1672     GCCTTGGCGCAGCGCCCTGTGTGGCCATCAGAGATCCGTCGACGACTGTGGACACCAT
      1682 1687      1699 1704
1732     CATCCGAGGCTGTGCTATCGA
...
1766     CTCAGTCAACGTGACACGCGTGTATGGGACACCGCGCAGCGCGGCGGTGATCGCCCTG
      1816 1821
1826     CGGCGGCCAAGGTGTGTGACGCTGATTTCCATTACCCCAAGCA
...
-----
11 matches found in sequence:
aav4443: Mycobacterium tuberculosis antigen RDIF8 DNA.
(from "mycobacterieng.seq")
FOIG of: aav4443 check: 5767 from: 1 to: 1055
ID AAV44443 standard; DNA; 1055 BP.
XX
```

```

AC AAV44443;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen RDIF8 DNA.
DE
XX
XX Tuberculosis; infection; diagnosis; antigen; RDIF8; ss.
XX
XX Mycobacterium tuberculosis strain Erdman.
XX
XX Key Location/Qualifiers
XX CDS 1..270
XX /*tag= a
XX
XX M09816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX P-PSDB; AAW64374.
XX
XX WPI: 1998-251292/22.
XX P-PSDB; AAW64374.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and diagnosis of tuberculosis
XX
XX Claim 11a; page 199; 250pp; English.
XX
XX This DNA sequence codes for an antigenic portion (see AAW64374) of
XX Mycobacterium tuberculosis antigen RDIF8. It was isolated from a
XX M. tuberculosis strain Erdman genomic DNA expression library using
XX rabbit anti-sera raised against M. tuberculosis fractionated
XX proteins. Phage plaques expressing immunoreactive antigens were
XX purified. Recombinant RDIF8 was shown to stimulate cell
XX proliferation and interferon-gamma production in T cells isolated
XX from M. tuberculosis-immune donors. The invention relates to
XX compositions and methods for diagnosing tuberculosis. It provides
XX polypeptides (see AAW64291-W64379) comprising antigenic or immunogenic
XX portions of M. tuberculosis antigens, as well as DNA sequences
XX encoding such polypeptides, recombinant expression vectors and
XX transformed or transfected host cells. Also claimed are methods
XX and diagnostic kits for detecting M. tuberculosis infection in a
XX patient using these polypeptides, antibodies or oligonucleotide
XX probes and primers.
XX
XX Sequence 1055 BP; 191 A; 289 C; 375 G; 199 T; 1 other;
XX
AAV44443 Length: 1055 March 5, 2002 14:19 Type: N Check: 5767
Found using 'seq2-3' (pappu403.key)
1 CTGGCGTGCAGTGTACCGGCATATGAGTCGGCATTCATTTCGGCGCCCGCCGCGGA
1
1-----1-----1-----1-----1-----1-----1-----1-----1
61 CCGGTGCGCCCACTGATGAGCACCAGCGTCCGTCATATGCGAAGTGCCTGCGT
64 69 107 112
121 GCCCATGTGCTGCTGGGTTTCTTGAACGACGAGCCCTGCCGTATCGGCTCCCTACCCCCA
181 AACAAGCGCCAGTCAGGATCCGTCGCCGCGCGCGCATTCACAGCGGCGCATGTGCTAGC
184 189

```

```

...
514 GTTCCGCTCAGCTGGGAGTGGCTGTGCCACCCGATGATATGACCACCTGGGCGCTGCG
1-----1-----1-----1-----1-----1-----1-----1-----1
574 CCGGAGACGGCGCGGATGTCGATGTCAGCGCGCGAAGGGCGGACGACGAGCGCGCG
583 588
634 GGCATGACAGAGTGGGATGATGCGAGCGCTGGAACGAGTGGGTGCGGAGACCTGAA
1-----1-----1-----1-----1-----1-----1-----1-----1
694 CCGCGCTTTGAGTGCACAGGATGACAGACCGTGAATTCGCATTCCTCCGGCGCGCG
754 TAGGAGAGGGGGCGCAGACTGTCTTATTGACCACTGATCGCGCTCGGTGTTCCCG
1-----1-----1-----1-----1-----1-----1-----1-----1
814 CGCGCGGCTATGACAA 774 779
...
942 CGTTTGTAGTGCAACGCCAGACGGTGGAGAGCAGAGCGCTCCCGATGTGGCGCTCCG
1-----1-----1-----1-----1-----1-----1-----1-----1
992 997
1002 CAAACATCTCGGGGCGGCTGGAGTGGCATGGCCGAGGAGACTCGCTAGAC
1-----1-----1-----1-----1-----1-----1-----1-----1
1040 1045
4 matches found in sequence:
aav44444: Mycobacterium tuberculosis antigen RDIF10 DNA.
(from "mycobacterieng.seq")
TOIG of: aav44444 check: 8755 from: 1 to: 359
ID AAV44444 standard; DNA: 359 BP.
XX
XX AAV44444;
AC
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen RDIF10 DNA.
XX
XX Tuberculosis; infection; diagnosis; antigen; RDIF10; ss.
XX
XX Mycobacterium tuberculosis strain Erdman.
XX
XX Key Location/Qualifiers
XX CDS 3..359
XX /*tag= a
XX
XX M09816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
XX P-PSDB; AAW64375.
XX
XX WPI: 1998-251292/22.
XX P-PSDB; AAW64375.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX

```

PT and diagnosis of tuberculosis
 XX Claim 11a; Page 200; 250pp; English.
 XX
 CC This DNA sequence codes for an antigenic portion (see AAW64375) of
 CC Mycobacterium tuberculosis antigen RDIF10. It was isolated from a
 CC M. tuberculosis strain Erdman genomic DNA expression library using
 CC rabbit anti-sera raised against M. tuberculosis fractionated
 CC proteins. Phage plaques expressing immunoreactive antigens were
 CC purified. Recombinant RDIF10 was shown to stimulate cell
 CC proliferation and interferon-gamma production in T cells isolated
 CC from M. tuberculosis-immune donors. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising antigenic or immunogenic
 CC portions of M. tuberculosis antigens, as well as DNA sequences
 CC encoding such polypeptides, recombinant expression vectors and
 CC transformed or transfected host cells. Also claimed are methods
 CC and diagnostic kits for detecting M. tuberculosis infection in a
 CC patient using these polypeptides, antibodies or oligonucleotide
 CC probes and primers.
 XX
 SQ Sequence 359 BP; 64 A; 127 C; 113 G; 55 T; 0 other;
 AAV44444 Length: 359 March 5, 2002 14:19 Type: N Check: 8755 ..
 Found using 'seq2-3' (pappu403.key)
 ...
 29 GGCGCGCTCGACCGCACTGCGCGGTGTGTGCGGGCTGACCAACCGGATGCCGAAAC |-----|
 89 CATCCGAGATCACCCTCGCAATGATCCACCTCGCGCAGCTGCTCACCACCGACCGGGGCGG
 149 TGTGCGAGCGGCTGCATCCCTTGATAGCCGTCGCGCCCGACCGCAGGAAGTTGT |-----|
 157 162
 209 AGTA
 ...
 217 CCCACCACTGTTACCGGAGCGAGAAAGTTGAGGGTGAAGTCGCGCATGTGCGCGCGG |-----|
 267 272
 277 AGGTAGTTGACCCGGAAACCAAGATCCTCGGCAAGGTGCTCGGGCCCGGCACACGACA
 337 AACCCGACGCGCGGATAGGTACG |-----|
 342 347

 3 matches found in sequence:
 aav44445 : Mycobacterium tuberculosis antigen RDIF11 DNA.
 (from "Mycobacterium.seq")
 TOIG of: aav44445 check: 9229 from: 1 to: 350
 ID AAV44445 standard; DNA; 350 BP.
 XX
 AC AAV44445;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen RDIF11 DNA.
 XX
 KM Tuberculosis; infection; diagnosis; antigen; RDIF11; ss.
 XX
 OS Mycobacterium tuberculosis strain Erdman.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..350
 FT /*tag= a

XX
 PN MO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97MO-US18214.
 XX
 PR 13-MAR-1997; 97US-0818111.
 PR 11-OCT-1996; 96US-0729622.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skelky YAM, Twardzik DR, Vedvick TS;
 DR WPI; 1998-251292/22.
 DR P-PSDB; AAW64376.
 XX
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis
 PS Claim 11a; Page 200; 250pp; English.
 XX
 CC This DNA sequence codes for an antigenic portion (see AAW64376) of
 CC Mycobacterium tuberculosis antigen RDIF11. It was isolated from a
 CC M. tuberculosis strain Erdman genomic DNA expression library using
 CC rabbit anti-sera raised against M. tuberculosis fractionated
 CC proteins. Phage plaques expressing immunoreactive antigens were
 CC purified. Recombinant RDIF11 was shown to stimulate cell
 CC proliferation and interferon-gamma production in T cells isolated
 CC from M. tuberculosis-immune donors. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising antigenic or immunogenic
 CC portions of M. tuberculosis antigens, as well as DNA sequences
 CC encoding such polypeptides, recombinant expression vectors and
 CC transformed or transfected host cells. Also claimed are methods
 CC and diagnostic kits for detecting M. tuberculosis infection in a
 CC patient using these polypeptides, antibodies or oligonucleotide
 CC probes and primers.
 XX
 SQ Sequence 350 BP; 64 A; 110 C; 106 G; 70 T; 0 other;
 AAV44445 Length: 350 March 5, 2002 14:19 Type: N Check: 9229 ..
 Found using 'seq2-3' (pappu403.key)
 ...
 65 CATCTTGGCAGATTCCCGCCAGAGCAAAACAGCCGCTAGTCTAGTCCGAGTGGCCGCA |-----|
 115 120
 125 AAGTTCCTCGAATACCTCCGTACCGGAGCGCCAAACCGGGTCTCTCGTAAGCTGCG |-----|
 152 157
 185 CGAACCACTTGAGGTTCCGGGACTCCTTGACGTCACAGACGATTCGTTGAGTGGCTGAT |-----|
 213 218
 245 CGGTTGCGCGCGCTGGGCGCAATC
 ...

 15 matches found in sequence:
 aav44446 : Mycobacterium tuberculosis antigen RDIF12 5' DNA.
 (from "Mycobacterium.seq")
 TOIG of: aav44446 check: 4816 from: 1 to: 811
 ID AAV44446 standard; DNA; 811 BP.
 XX
 AC AAV44446;

PT and diagnosis of tuberculosis
 XX
 PS Claim 11a; Page 207; 250pp; English.
 XX
 CC This is the 3' sequence of DNA coding for an antigenic portion of
 CC Mycobacterium tuberculosis antigen RD1F12; a 5' portion of RD1F12
 CC DNA is provided in AAV44446. RD1F12 DNA was isolated from a M.
 CC tuberculosis strain Erdman genomic DNA expression library using
 CC rabbit anti-sera raised against M. tuberculosis fractionated
 CC proteins. Phage plaques expressing immunoreactive antigens were
 CC purified. Recombinant RD1F12 was shown to stimulate cell
 CC proliferation and interferon-gamma production in T cells isolated
 CC from M. tuberculosis-immune donors. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAM64291-W64379) comprising antigenic or immunogenic
 CC portions of M. tuberculosis antigens, as well as DNA sequences
 CC encoding such polypeptides, recombinant expression vectors and
 CC transformed or transfected host cells. Also claimed are methods
 CC and diagnostic kits for detecting M. tuberculosis infection in a
 CC patient using these polypeptides, antibodies or oligonucleotide
 CC probes and primers.
 XX
 SQ Sequence 966 BP; 199 A; 297 C; 285 G; 185 T; 0 other;
 AAV44447 Length: 966 March 5, 2002 14:19 Type: N Check: 4267 ..
 Found using 'seq2-3' (pappu403.key)

48 GTCGAGATATCGGACTTGTGTGTCGCCGGTGGCGGATAGAGACCTGTGCGGCTGTC
 98 103
 108 |-----|
 108 AGCGTACCCGTTGTCGGAGCGCGCAACCCATCTTCAACGTAAGCTGTGCTCACACA
 108 113 126 131
 168 |-----|
 180 185
 228 |-----|
 228 GCGGACATCAAGGCACTACTACCTTACATCGAGCCATTGTGACGGGTGATCGCGCTTC
 279 284
 288 |-----|
 318 323
 348 |-----|
 348 CACGAGGTGCAGCCCAAGCAACGCG
 403 |-----|
 403 CGTGTGGCGAGTGTGCACTACCGGATCTTCCCGGGCCTTACGAGTCCGGGAGCGCC
 453 458
 463 |-----|
 470 475 480 485 489 494
 523 |-----|
 523 TCGAGGTGTTGAGATTGCTCC
 555 |-----|
 555 GTGACCCCGATGATGTTGCAGCCGAGCTACTTTTGGAGCCTGATTCGAAAGTGTGAA
 605 610
 615 |-----|
 615 TCGTATGTGCTGCTGCTAAGACATCGCCGAGACCGACCGAGTGGGATCGTGGATCGC

675 |-----|
 675 CCCACCGCGCGGTGATTCAGGAAAAATAAGAGCCGCTATCCACATTCGGCGTGCAGCTC
 723 728
 735 |-----|
 735 GGCTACCAACAACGGTAGACGATCGAGACATTCCGAGCTGAAGTGGCGCTATGAA
 753 758
 795 |-----|
 795 GCGGCTGTGCGGATTAATCAACGCAAAATACCTTACTCATGCATCGGCTGCTAC
 843 848
 855 |-----|
 855 CCGATGCGAGCTTTTTCACAGCTCCACCGCTGCGCGGAGCTCAATGGGATGCAT
 862 867
 915 |-----|
 915 CCCACCGGTTCGCGAAGCGGTTCCGGCGGTGCTCATGCTTCATCT
 954 959

 1 match found in sequence:
 aav68325; Murine STIF cDNA.
 (from "Mycobacterieng.seq")
 TOIG of: aav68325 check: 6052 from: 1 to: 1018

ID AAV68325 standard; cDNA; 1018 BP.
 XX
 AC AAV68325;
 XX
 DT 05-MAR-1999 (first entry)
 XX
 DE Murine STIF cDNA.
 XX
 KW STIF: TH2 cell subpopulation; TH2-specific gene; immune disorder;
 KW T-lymphocyte related disorder; cell proliferation; cancer; psoriasis;
 KW chronic inflammatory disease; Crohn's disease; arthritis; Lyme disease;
 KW insulin-dependent diabetes; organ-specific autoimmunity; Grave's disease;
 KW multiple sclerosis; Hashimoto's thyroiditis; dermatitis; sarcoidosis;
 KW asthma; allergy; eosinophilia; conjunctivitis; glomerular nephritis;
 KW leishmaniasis; viral infection; HIV; bacterial infection; tuberculosis;
 KW lepromatous leprosy; ss.
 XX
 OS Mus spretus.
 XX
 FH Key Location/Qualifiers
 FT CDS 73..735
 FT /*tag= a
 XX
 PN WO9846638-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 17-APR-1998; 98WO-US07616.
 XX
 PR 25-JUN-1997; 97US-0884077.
 PR 17-APR-1997; 97US-0841901.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Gu W, Lehar S, Levinson D;
 XX
 DR WPI: 1998-594560/50.
 DR P-PSDB: AAW81138.
 XX
 PT New isolated TH2-specific gene, STIF - used to develop products for
 PT treating an immune disorder, e.g. a T-lymphocyte related disorder or
 PT a proliferative disorder such as cancer
 XX
 PS Claim 2; Fig 1; 125pp; English.

CC The present sequence encodes murine STIF protein, which is a CD4+ helper
CC T lymphocyte (TH2) cell subpopulation protein. The STIF protein is
CC differentially expressed within the TH2 cell subpopulation. The STIF
CC polypeptides can stimulate or inhibit cytokine production by cells. The
CC STIF gene can be used for identifying compounds which modulate STIF
CC activity. The compounds can be used for treating an immune disorder,
CC e.g. a T-lymphocyte related disorder or a proliferative disorder, e.g.
CC cancer. They can be used for treating e.g. chronic inflammatory diseases
CC and disorders, such as Crohn's disease, reactive arthritis, Lyme disease,
CC insulin-dependent diabetes, organ-specific autoimmunity, multiple
CC sclerosis, Hashimoto's thyroiditis, Grave's disease, contact dermatitis,
CC sarcoidosis, asthma and allergy, including allergic rhinitis,
CC gastrointestinal allergies, including food allergies, eosinophilia,
CC conjunctivitis, glomerular nephritis, certain pathogen susceptibilities
CC such as helminthic (e.g. leishmaniasis) and certain viral infections,
CC including HIV, and bacterial infections, including tuberculosis,
CC lepromatous leprosy, atopic, IGE-mediated allergic conditions, including
CC pathogen susceptibilities and chronic inflammatory disease, or psoriasis.
CC The products can also be used for detection, diagnosis and drug
CC screening.
CC
CC
SQ Sequence 1018 BP; 249 A; 281 C; 223 G; 265 T; 0 other:
AAV68325 Length: 1018 March 5, 2002 14:19 Type: N Check: 6052 ..
Found using 'seq2-3' (pappu403.key)
...
614 ACCAGCGGTTTGTGCTTCGCGAGACATTCAAACAGTTGGATACAGAGATCCCTTTGG |-----|
ACCAAGCCTTTGGGGAAGTGACATTCCTCGACCTGATGCACAGA 664 669
674 TGAAGCCTTTGGGGAAGTGACATTCCTCGACCTGATGCACAGA
...
4 matches found in sequence:
aaV68326; Human STIF cDNA.
(from "mycobacterieng.seq")
TOIG of: aaV68326 check: 6041 from: 1 to: 1568
ID AAV68326 standard; cDNA; 1568 BP.
XX
XX AAV68326;
AC
XX
DT 05-MAR-1999 (first entry)
XX
XX Human STIF cDNA.
DE
XX
XX STIF: TH2 cell subpopulation; TH2-specific gene; immune disorder;
KW T-lymphocyte related disorder; cell proliferation; cancer; psoriasis;
KW chronic inflammatory disease; Crohn's disease; arthritis; Lyme disease;
KW insulin-dependent diabetes; organ-specific autoimmunity; Grave's disease;
KW multiple sclerosis; Hashimoto's thyroiditis; dermatitis; sarcoidosis;
KW asthma; allergy; eosinophilia; conjunctivitis; glomerular nephritis;
KW leishmaniasis; viral infection; HIV; bacterial infection; tuberculosis;
KW lepromatous leprosy; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 209..748
FT /*tag= a
FT
XX
XX WO9846638-A1.
PN
XX
PD 22-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07616.
XX
XX 25-JUN-1997; 97US-0884077.
PR 17-APR-1997; 97US-0841901.
PR

XX
PA (MILL-) MILLENNIUM PHARM INC.
XX Gu W, Lohar S, Levinson D;
XX
XX
XX WPI: 1998-594560/50.
DR P-PSDB: AAM81139.
XX
XX
XX New isolated TH2-specific gene, STIF - used to develop products for
PT treating an immune disorder, e.g. a T-lymphocyte related disorder or
PT a proliferative disorder such as cancer
PS
XX Claim 2; Fig 2; 125pp; English.
XX
XX The present sequence encodes human STIF protein, which is a CD4+ helper
CC T lymphocyte (TH2) cell subpopulation protein. The STIF protein is
CC differentially expressed within the TH2 cell subpopulation. The STIF
CC polypeptides can stimulate or inhibit cytokine production by cells. The
CC STIF gene can be used for identifying compounds which modulate STIF
CC activity. The compounds can be used for treating an immune disorder,
CC e.g. a T-lymphocyte related disorder or a proliferative disorder, e.g.
CC cancer. They can be used for treating e.g. chronic inflammatory diseases
CC and disorders, such as Crohn's disease, reactive arthritis, Lyme disease,
CC insulin-dependent diabetes, organ-specific autoimmunity, multiple
CC sclerosis, Hashimoto's thyroiditis, Grave's disease, contact dermatitis,
CC sarcoidosis, asthma and allergy, including allergic rhinitis,
CC gastrointestinal allergies, including food allergies, eosinophilia,
CC conjunctivitis, glomerular nephritis, certain pathogen susceptibilities
CC such as helminthic (e.g. leishmaniasis) and certain viral infections,
CC including HIV, and bacterial infections, including tuberculosis,
CC lepromatous leprosy, atopic, IGE-mediated allergic conditions, including
CC pathogen susceptibilities and chronic inflammatory disease, or psoriasis.
CC The products can also be used for detection, diagnosis and drug
CC screening.
CC
CC
SQ Sequence 1568 BP; 401 A; 394 C; 349 G; 424 T; 0 other:
AAV68326 Length: 1568 March 5, 2002 14:19 Type: N Check: 6041 ..
Found using 'seq2-3' (pappu403.key)
...
143 CTCAGCATCTTACAGAGACAGACAGACACCTTTGCCCCCTTGTGCGGACAGCC |-----|
193 198
203 TCTCAATGACAGATGTTGTGCTCCCTTGCTGGGTTTACCTGC
...
372 CTCAGATACATCAGAGATGCCCGGCTGTGACAGAGAGATTCTGCAGAACGCTTCGG |-----|
422 427
432 ATGCTGAGAGCTGTACTTGTGCACACCCCTGCGAGTTCTACTT
...
857 GGGTCCCATCTTGGCCCCAGAGATTATTGTCAAGAGACTCTTTTAAGCAGCCGCTG |-----|
907 912
917 ACAGTCAGGAGAGTGCCCTCTGATGCTGTGAAGAGTCTACAGAGA
...
1254 GCAGGACAGAGCATTTGGGGTGGGGGTAAAGTCATCTGTTGAAGAAAGTAACGATAAA |-----|
1304 1309
1314 TGTGATTTAAGTGCCAGCAGCAAGAGATCTCTCAATAACATTT
...

2 matches found in sequence:

aav68327 : Human STIF CDNA with homology to MDA-7.

(from "mycobacterng.seq")

TOIG of: aav68327 check: 8505 from: 1 to: 621

ID AAV68327 standard; CDNA: 621 BP.
 AC AAV68327;
 XX
 XX
 DT 05-MAR-1999 (first entry)
 XX
 XX
 DE Human STIF CDNA with homology to MDA-7.
 KW STIF, TH2 cell subpopulation; TH2-specific gene; immune disorder;
 KW T-lymphocyte related disorder; cell proliferation; cancer; psoriasis;
 KW chronic inflammatory disease; Crohn's disease; arthritis; Lyme disease;
 KW insulin-dependent diabetes; organ-specific autoimmunity; Grave's disease;
 KW multiple sclerosis; Hashimoto's thyroiditis; dermatitis; sarcoidosis;
 KW asthma; allergy; eosinophilia; conjunctivitis; glomerular nephritis;
 KW leishmaniasis; viral infection; HIV; bacterial infection; tuberculosis;
 KW lepromatous leprosy; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..621
 FT /tag= a
 FT
 PN WO9846638-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 17-APR-1998; 98WO-US07616.
 XX
 PR 25-JUN-1997; 97US-0884077.
 PR 17-APR-1997; 97US-0841901.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA Gu W, Lehar S, Levinson D;
 PI
 DR WPI: 1998-594560/50.
 DR P-PSDB: AAW81140.
 XX
 PT New isolated TH2-specific gene, STIF - used to develop products for
 PT treating an immune disorder, e.g. a T-lymphocyte related disorder or
 PT a proliferative disorder such as cancer
 PS
 PS Disclosure: Fig 3; 125pp; English.
 XX
 CC The present sequence encodes human STIF protein which has homology to
 CC MDA-7. STIF is a CD4+ helper T lymphocyte (TH2) cell subpopulation
 CC protein. The STIF protein is differentially expressed within the TH2
 CC cell subpopulation. The STIF polypeptides can stimulate or inhibit
 CC cytokine production by cells. The STIF gene can be used for identifying
 CC compounds which modulate STIF activity. The compounds can be used for
 CC treating an immune disorder, e.g. a T-lymphocyte related disorder or a
 CC proliferative disorder, e.g. cancer. They can be used for treating e.g.
 CC chronic inflammatory diseases and disorders, such as Crohn's disease,
 CC reactive arthritis, Lyme disease, insulin-dependent diabetes, organ-
 CC specific autoimmunity, multiple sclerosis, Hashimoto's thyroiditis,
 CC Grave's disease, contact dermatitis, sarcoidosis, asthma and allergy,
 CC including allergic rhinitis, gastrointestinal allergies, including food
 CC allergies, eosinophilia, conjunctivitis, glomerular nephritis, certain
 CC pathogen susceptibilities such as helminthic (e.g. leishmaniasis) and
 CC certain viral infections, including HIV, and bacterial infections,
 CC including tuberculosis, lepromatous leprosy, atopic, IGE-mediated
 CC allergic conditions, including pathogen susceptibilities and chronic
 CC inflammatory disease, or psoriasis. The products can also be used for
 CC detection, diagnosis and drug screening.
 CC
 CC

SQ Sequence 621 BP; 154 A; 161 C; 153 G; 153 T; 0 other;

AAV68327 Length: 621 March 5, 2002 14:19 Type: N Check: 8505 ...
 Found using 'seq-3' (pappu403.key)

19 CTCGAAGCCTGTGACTTTAGCCAGACAGACCCTTCTGCCCTTCTGCGGACAGCC
 69 74
 79 TCTCAATGCGAGATGGTTGTGCTCCTTGCCTGGGTTTACCTGC
 248 CTCGAGTATACATCAGAGTCCCGGCTGCTGCAGCAGAGTTTCAGACAGTTTCGG
 298 303
 308 ATGCTGAGAGCTGTACTTGTCCACACCCTGCTGAGTTTACTT
 ...
 2 matches found in sequence:
 aav68333 : Murine STIF oligonucleotide #2.
 (from "mycobacterng.seq")
 TOIG of: aav68333 check: 3635 from: 1 to: 64
 ID AAV68333 standard; DNA: 64 BP.
 XX
 AC AAV68333;
 XX
 DT 05-MAR-1999 (first entry)
 XX
 DE Murine STIF oligonucleotide #2.
 XX
 KW STIF, TH2 cell subpopulation; TH2-specific gene; immune disorder;
 KW T-lymphocyte related disorder; cell proliferation; cancer; psoriasis;
 KW chronic inflammatory disease; Crohn's disease; arthritis; Lyme disease;
 KW insulin-dependent diabetes; organ-specific autoimmunity; Grave's disease;
 KW multiple sclerosis; Hashimoto's thyroiditis; dermatitis; sarcoidosis;
 KW asthma; allergy; eosinophilia; conjunctivitis; glomerular nephritis;
 KW leishmaniasis; viral infection; HIV; bacterial infection; tuberculosis;
 KW lepromatous leprosy; PCR primer; ss.
 KW
 XX
 OS Synthetic.
 OS Mus spretus.
 OS
 PN WO9846638-A1.
 XX
 PD 22-OCT-1998.
 XX
 PF 17-APR-1998; 98WO-US07616.
 XX
 PR 25-JUN-1997; 97US-0884077.
 PR 17-APR-1997; 97US-0841901.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Gu W, Lehar S, Levinson D;
 DR WPI: 1998-594560/50.
 DR
 XX
 PT New isolated TH2-specific gene, STIF - used to develop products for
 PT treating an immune disorder, e.g. a T-lymphocyte related disorder or
 PT a proliferative disorder such as cancer
 PS
 PS Example: Page 103; 125pp; English.
 XX
 CC The present sequence represents an oligonucleotide used in an example
 CC from the present invention. The present invention describes STIF
 CC which is a CD4+ helper T lymphocyte (TH2) cell subpopulation protein.
 CC

CC The STIF protein is differentially expressed within the TH2 cell
 CC subpopulation. The STIF polypeptides can stimulate or inhibit cytokine
 CC production by cells. The STIF gene can be used for identifying compounds
 CC which modulate STIF activity. The compounds can be used for treating an
 CC immune disorder, e.g. a T-lymphocyte related disorder or a proliferative
 CC disorder, e.g. cancer. They can be used for treating e.g. chronic
 CC inflammatory diseases and disorders, such as Crohn's disease, reactive
 CC arthritis, Lyme disease, insulin-dependent diabetes, organ-specific
 CC autoimmunity, multiple sclerosis, Hashimoto's thyroiditis, Grave's
 CC disease, contact dermatitis, sarcoidosis, asthma and allergy, including
 CC allergic rhinitis, gastrointestinal allergies, including food allergies,
 CC eosinophilia, conjunctivitis, glomerular nephritis, certain pathogen
 CC susceptibilities such as helminthic (e.g. leishmaniasis) and certain
 CC viral infections, including HIV, and bacterial infections, including
 CC tuberculosis, lepromatous leprosy, atopic, IgE-mediated allergic
 CC conditions, including pathogen susceptibilities and chronic inflammatory
 CC disease, or psoriasis. The products can also be used for detection,
 CC diagnosis and drug screening.
 CC

SQ Sequence 64 BP; 12 A; 16 C; 15 G; 21 T; 0 other;

AAV68333 Length: 64 March 5, 2002 14:19 Type: N Check: 3635 ..
 Found using 'seq2-3' (pappu403.key)

1 GCTCTAGACGTCGAGTCACCTTGTCTCGTCGCTTGTAGTCGAGATGCTAGCAATTTCGCG

24 29 32

61 ATCC

3 matches found in sequence:

aaav68335; Interleukin 10 PCR 3' primer.
 (from "mycobacterieng.seq")
 TOIG of: aaav68335 check: 8633 from: 1 to: 67

ID AAV68335 standard; DNA; 67 BP.

AC AAV68335;

DT 05-MAR-1999 (first entry)

DE Interleukin 10 PCR 3' primer.

KW STIF; TH2 cell subpopulation; TH2-specific gene; immune disorder;
 KW T-lymphocyte related disorder; cell proliferation; cancer; psoriasis;
 KW chronic inflammatory disease; Crohn's disease; arthritis; Lyme disease;
 KW insulin-dependent diabetes; organ-specific autoimmunity; Grave's disease;
 KW multiple sclerosis; Hashimoto's thyroiditis; dermatitis; sarcoidosis;
 KW asthma; allergy; eosinophilia; conjunctivitis; glomerular nephritis;
 KW leishmaniasis; viral infection; HIV; bacterial infection; tuberculosis;
 KW lepromatous leprosy; PCR primer; ss.

OS Synthetic.

PN WO9846638-A1.

PD 22-OCT-1998.

PF 17-APR-1998; 98WO-US07616.

PR 25-JUN-1997; 97US-0884077.

PR 17-APR-1997; 97US-0841901.

PA (MILL-) MILLENNIUM PHARM INC.

PI Gu W, Lehar S, Levinson D;

DR WPI; 1998-594560/50.

XX New isolated TH2-specific gene, STIF - used to develop products for

PT treating an immune disorder, e.g. a T-lymphocyte related disorder or
 PT a proliferative disorder such as cancer

PS Example; Page 103; 125pp; English.

CC The present sequence represents a PCR primer used in an example from the
 CC present invention. The present invention describes a STIF protein,
 CC which is a CD4+ helper T lymphocyte (TH2) cell subpopulation protein.
 CC The STIF protein is differentially expressed within the TH2 cell
 CC subpopulation. The STIF polypeptides can stimulate or inhibit cytokine
 CC production by cells. The STIF gene can be used for identifying compounds
 CC which modulate STIF activity. The compounds can be used for treating an
 CC immune disorder, e.g. a T-lymphocyte related disorder or a proliferative
 CC disorder, e.g. cancer. They can be used for treating e.g. chronic
 CC inflammatory diseases and disorders, such as Crohn's disease, reactive
 CC arthritis, Lyme disease, insulin-dependent diabetes, organ-specific
 CC autoimmunity, multiple sclerosis, Hashimoto's thyroiditis, Grave's
 CC disease, contact dermatitis, sarcoidosis, asthma and allergy, including
 CC allergic rhinitis, gastrointestinal allergies, including food allergies,
 CC eosinophilia, conjunctivitis, glomerular nephritis, certain pathogen
 CC susceptibilities such as helminthic (e.g. leishmaniasis) and certain
 CC viral infections, including HIV, and bacterial infections, including
 CC tuberculosis, lepromatous leprosy, atopic, IgE-mediated allergic
 CC conditions, including pathogen susceptibilities and chronic inflammatory
 CC disease, or psoriasis. The products can also be used for detection,
 CC diagnosis and drug screening.
 CC

SQ Sequence 67 BP; 11 A; 16 C; 13 G; 27 T; 0 other;

AAV68335 Length: 67 March 5, 2002 14:19 Type: N Check: 8633 ..
 Found using 'seq2-3' (pappu403.key)

1 GCTCTAGACGTCGAGTCACCTTGTCTCGTCGCTTGTAGTCGAGATGCTAGCAATTTCGATC

24 29 32 39 44

61 ATGTATG

7 matches found in sequence:

aaav69310; E. coli ori genomic DNA.
 (from "mycobacterieng.seq")
 TOIG of: aaav69310 check: 9353 from: 1 to: 695

ID AAV69310 standard; DNA; 695 BP.

AC AAV69310;

DT 01-FEB-1999 (first entry)

DE E. coli ori genomic DNA.

KW Origin of replication; vaccine; cellular immunity; Th-1 cell; cancer;
 KW cytotoxic T lymphocyte; treatment; prevention; disease; autoimmune;
 KW intracellular pathogen; allergy; bovine spongiform encephalitis; BSE;
 KW macrophage; immunogen; adjuvant; ds.
 KW Escherichia coli.

OS Escherichia coli.

PN WO9844096-A2.

PD 08-OCT-1998.

PF 27-MAR-1998; 98WO-US06056.

PR 28-MAR-1997; 97US-0042849.

XX


```

XX (CYTO) CYTOCLONAL PHARM INC.
XX
XX Labidi AH:
XX
XX WPI: 1998-568277/48.
XX
XX Recombinant non-pathogenic Mycobacterium as vaccines providing long
XX term cellular immunity - useful against intracellular pathogens,
XX cancer and autoimmune disease, and are retained in host macrophages
XX
XX Claim 60: Fig 1; 120bp: English.
XX
XX This genomic DNA sequence contains the Escherichia coli origin of
XX replication, ori. This sequence is used in a method which results in the
XX formation of Mycobacterium recombinant vaccines from compositions that
XX provide a continuous source of protein to an animal and stimulate
XX cellular immunity. Such compositions are used to stimulate cellular
XX immunity (by inducing Th-1 cells or cytotoxic T lymphocytes),
XX particularly as vaccines (live or dead) for treating and preventing
XX diseases caused by intracellular pathogens (bacteria, viruses, rickettsia
XX or protozoa), and also cancer, autoimmune diseases, allergy and bovine
XX spongiform encephalitis, in humans or animals. The vaccines are
XX administered by injection, orally and nasally. These compositions provide
XX consistent and long-lasting immunity. Transformed cells used in the
XX method are retained within macrophages, blocking the killing mechanism
XX but producing protective immunogen which is processed and presented by
XX the macrophage. Several immunogens may be included in the same vaccine
XX and the Mycobacterium cells serve as adjuvant.
XX
XX Sequence 695 BP; 162 A; 188 C; 176 G; 169 T; 0 other;
XX
AAV69310 Length: 695 March 5, 2002 14:17 Type: N Check: 9353 ..
Found using 'seq2-3' (pappu403.key)
1 GTTTTTCATAGGCTCCGCCCCCTGACGACATCACAAAATGCGACCTCAAGTCAGAG
|-----| |-----|
61 GTGGCGAAACCCGACAGAGATATAAGATACACAGGCTTCCCCCTGAAAGCTCCCTGCT
|-----| 94 99
121 GCGCTCTCCTGTTCCGACCCCTGCGCTTACCGAGATACCTGTGCGCTTCTCCCTTCGGG
|-----| |-----|
181 AACCGTGGGCTTCTTCATAGTCACAGCCTTAGGTATCTAGTTGGTGTAGGCTGCTCG
187 192 233 238
241 CTCGAAGCTGGGCTGTGTGCGACAACCCCGTTCAAGCCGACCGCTGCGCTTATCCGG
|-----| |-----|
301 TAACCTATGCGCTTGAGTCCAAACCCGTAAGACACAGACTTATCGGCACTGGCAGCAGCCAC
306 311 340 345
361 TGGTAACAGGATTAGCAGACGACGAGTATGTAGGCG
|-----|
573 GCAGAAAAAAGATCTCAAGAAGATCCTTGATCTTTCTACGGGGCTGACGCTACT
|-----|
623 GGAACGAAACTCACGTTAAGGATTTTGGTCATGAGATTATCAAA
|-----|
4 matches found in sequence:
aav69311 : E. coli kanamycin genomic DNA.
(from "mycobacterin.seq")
FOIG of: aav69311 check: 5171 from: 1 to: 932

```

```

ID AAV69311 standard; DNA; 932 BP.
XX AAV69311;
XX 01-FEB-1999 (first entry)
DE E. coli kanamycin genomic DNA.
KW Origin of replication; vaccine; cellular immunity; Th-1 cell; cancer;
KW cytotoxic T lymphocyte; treatment; prevention; disease; autoimmune;
KW intracellular pathogen; allergy; bovine spongiform encephalitis; BSE;
KW macrophage; immunogen; adjuvant; ds.
XX Escherichia coli.
OS
XX
PS Key Location/Qualifiers
FH -35_signal 2..7
FT /*tag= a
FT -10_signal 26..31
FT /*tag= b
FT RBS 92..99
FT /*tag= c
FT CDS 103..932
FT /*tag= d
FT /product= "Kanamycin"
XX
XX MO9844096-A2.
XX
XX 31-OCT-1998.
XX
XX 27-MAR-1998; 98WO-US06056.
XX
XX 28-MAR-1997; 97US-0042849.
XX
XX (CYTO-) CYTOCLONAL PHARM INC.
XX
XX Labidi AH;
XX
XX WPI; 1998-568277/48.
XX
XX Recombinant non-pathogenic Mycobacterium as vaccines providing long
PT term cellular immunity - useful against intracellular pathogens,
PT cancer and autoimmune disease, and are retained in host macrophages
XX
XX Claim 6; Fig 2; 120p; English.
XX
XX This genomic DNA sequence contains the Escherichia coli kanamycin gene.
CC This sequence is used in a method which results in the formation of
CC Mycobacterium recombinant vaccines from compositions that provide a
CC continuous source of protein to an animal and stimulate cellular
CC immunity. Such compositions are used to stimulate cellular immunity (by
CC inducing Th-1 cells or cytotoxic T lymphocytes), particularly as vaccines
CC (live or dead) for treating and preventing diseases caused by
CC intracellular pathogens (bacteria, viruses, rickettsia or protozoa), and
CC also cancer, autoimmune diseases, allergy and bovine spongiform
CC encephalitis, in humans or animals. The vaccines are administered by
CC injection, orally and nasally. These compositions provide consistent and
CC long-lasting immunity. Transformed cells used in the method are retained
CC within macrophages, blocking the killing mechanism but producing
CC protective immunogen which is processed and presented by the macrophage.
CC Several immunogens may be included in the same vaccine and the
CC Mycobacterium cells serve as adjuvant.
XX
XX Sequence 932 BP; 255 A; 179 C; 215 G; 283 T; 0 other;
AAV69311 Length: 932 March 5, 2002 14:17 Type: N Check: 5171
Found using 'seq2-3' (pappu403.key)

```

133 TCGAGCGCGCATTAATTCACATCATGCTGATTTATATGGGT
 ...
 230 TGTATGGGAACCCGATGCCGACAGTGTTCGTAACATGCGCAAGTAGCGTTCGCA
 280 285
 290 ATGATGTTACAGATGAGATGTCAGACTAACTGCTGACGGAATT
 ...
 479 TGTTCCTGCGCGCGTTCGATTCGATTCCTGTTGTAATTCCTTTAACGCGATCCGCG
 529 534
 539 TATTCGTCGTCGTCAGGCGCATACGAAATGAAATGCGTTGGT
 ...
 629 AACAGTCTGGAAAGAAATGCATAGCTTTTGCATTCCTACCGGATTCAGTCCGCACTC
 679 684
 689 ATGCTGATTCTCATTGATTAACCTTATTTTGAAGGGAATT
 ...
 25 matches found in sequence:
 aav69312; M. fortuitum plasmid pAL 5000 ori DNA.
 (from "mycobacteriophage")
 TOIG of: aav69312 check: 5740 from: 1 to: 1463

ID AAV69312 standard; DNA; 1463 BP.
 XX
 AC AAV69312:
 XX
 DT 01-FEB-1999 (first entry)
 XX
 DE M. fortuitum plasmid pAL 5000 ori DNA.
 XX
 KW Origin of replication; vaccine; cellular immunity; Th-1 cell; cancer;
 KW cytotoxic T lymphocyte; treatment; prevention; disease; autoimmune;
 KW intracellular pathogen; allergy; bovine spongiform encephalitis; BSE;
 KW macrophage; immunogen; adjuvant; ds.
 XX
 OS Mycobacterium fortuitum.
 XX
 PN WO9844096-A2.
 XX
 PD 31-OCT-1998.
 XX
 PF 27-MAR-1998; 98WO-US06056.
 XX
 PR 28-MAR-1997; 97US-0042849.
 XX
 PA (CYTO-) CYTOCLONAL PHARM INC.
 XX
 PI LabId: AH;
 XX
 DR WPI; 1998-568277/48.
 XX
 PT Recombinant non-pathogenic Mycobacterium as vaccines providing long
 PT term cellular immunity - useful against intracellular pathogens,
 PT cancer and autoimmune disease, and are retained in host macrophages
 XX
 PS Claim 29; Fig 3A; 120pp; English.
 XX
 CC This genomic DNA sequence contains is the Mycobacterium fortuitum
 CC plasmid pAL 5000 origin of replication, ori. This sequence is used in a
 CC method which results in the formation of Mycobacterium recombinant

CC vaccines from compositions that provide a continuous source of protein to
 CC an animal and stimulate cellular immunity. Such compositions are used to
 CC stimulate cellular immunity (by inducing Th-1 cells or cytotoxic T
 CC lymphocytes), particularly as vaccines (live or dead) for treating and
 CC preventing diseases caused by intracellular pathogens (bacteria, viruses,
 CC rickettsia or protozoa), and also cancer, autoimmune diseases, allergy
 CC and bovine spongiform encephalitis, in humans or animals. The vaccines
 CC are administered by injection, orally and nasally. These compositions
 CC provide consistent and long-lasting immunity. Transformed cells used in
 CC the method are retained within macrophages, blocking the killing
 CC mechanism but producing protective immunogen which is processed and
 CC presented by the macrophage. Several immunogens may be included in the
 CC same vaccine and the Mycobacterium cells serve as adjuvant.
 XX
 S0 Sequence 1463 BP; 294 A; 487 C; 461 G; 221 T; 0 other;
 AAV69312 Length: 1463 March 5, 2002 14:17 Type: N Check: 5740 ..
 Found using 'seq2-3' (pappu403.key)

139 CGGACATACACCGGTGATTAATGCTGCTACTACCAAGCGTGAGCCACCTCCGCGACG
 189 194
 199 AATTGAGCAGCTCTGGCTGCGTACTGCGCGTGGAGGAGCATCTGCTCGAGGGA
 237 242 240 245
 259 TCTACCGCAAGAGCCGCGCTGCGCCCTAGAGCGCGCGGTACATGAGGCGAACCACAG
 317
 319 CGGTGGCAAAACCTGCTGCTGCTGAGCATGACATCCAGACGACGCTCCGACGCTCA
 322 362 367 371 376
 379 GCGCGCGGGGTCCCATCCGCTGCCCAAGCGATGCTGGCAATCGCGCCCAAGCGCACG
 383
 439 CACACGAGTGTGGGCACTCAAGCCCTGTTCCACGACCGAGATPAGCGCGGCTAAGC
 459 464
 499 CGCTCGCATACATGGCGCGGTGCGCGCAAGGCTTGGCGCGCGCGCGACGCGCA
 549 554
 559 GTTACTCAGGCTTCATGACCAAAAACCCGCGCATCTGCTGGGAACGAATGCTCC
 594 599
 619 ACTCAGATCTCTACACATCAGCCACATCGAGGCGGCTTGGCGGCAATGCGACCGC
 679 CGGCGTGGCGTCACAGACCAAGCAAGGCGCTCGGAGCGGCTAGGCGGGAATTGCG
 685 690 716 721
 739 CACTGTTGATTCCTGCTGAGGTGTGGGCTATGCTCCGCGCTCATCGATCTACCTGC
 769 774
 799 CGACCGGAAGTGGACGACTCGCGCGCGATATATGCGGAGTCCACGCGGGAACG
 855
 859 -|-----|
 CCGAATTCCTGCAAGAGCTGTGTCGCGGACCGCTACCGGACAGAGGTCCGCGCA
 860 873 878 918

919 -----|-----|
TCGCCACAGCATTTGGCGATTCACAAACAGTGGCGATTTGGCGGAGGATCG
923 934 939
979 TGGCTACGAG
...
1056 ACAGTTGGCGGGCGGCAAGTCCGCGTCAGCCATGAGGCGATTGCTATGAGCGACGCT
|-----|
1116 ACAGCGACGCGTACAGCGAGGCTACACCGGACCGGACTGTCCGCAAAAAGCGGTGAC
1118 1123 1130 1135
1176 GCGCGGAAGCGCGCTGCAATACCGGACTATCCGACCGCCACGCTGCTGCGGCGC
1184 1189 1209 1214 1222
1236 AGGAACGCGAGAGTGGCTCGCCGAGCAGAGGCTGCAGCGCGCGCAAGCATCCGCCCTATC
|-----|
1296 ACAGCGACGAGGCGGCGCTATCGCGGAGGAAGAGCGTGGCGGACAGAGCGGCTC
1298 1303
1356 CCGTATAGCGACGCGCTATCGCGGAGGAAGAGCGTGGCGGACAGAGCGGCTC
1362 1367
1416 AA
...
25 matches found in sequence:
aav69313; M. fortuitum plasmid pAL 5000 ori DNA (derived by PCR analysis).
(from "mycobacteri3ng.seq")
TOIG of: aav69313 check: 5050 from: 1 to: 1382
ID AAV69313 standard; DNA: 1382 BP.
XX
AC AAV69313:
XX
DT 01-FEB-1999 (first entry)
XX
DE M. fortuitum plasmid pAL 5000 ori DNA (derived by PCR analysis).
XX
KW Origin of replication; vaccine; cellular immunity; Th-1 cell; cancer;
KW cytotoxic T lymphocyte; treatment; prevention; disease; autoimmune;
KW intracellular pathogen; allergy; bovine spongiform encephalitis; BSE;
KW macrophage; immunogen; adjuvant; ds.
XX
OS Mycobacterium fortuitum.
XX
PN WO9844096-A2.
XX
PD 31-OCT-1998.
XX
PF 27-MAR-1998; 98WO-US06056.
XX
PR 28-MAR-1997; 97US-0042849.
XX
PA (CYTO-) CYTOCLONAL PHARM INC.
XX
PI LabId: AH;
XX
DR WPI: 1998-568277/48.
XX
PT Recombinant non-pathogenic Mycobacterium as vaccines providing long
PT term cellular immunity - useful against intracellular pathogens,
PT cancer and autoimmune disease, and are retained in host macrophages
XX

PS Claim 30; Fig 3B; 120pp; English.
XX
CC This genomic DNA sequence contains is the Mycobacterium fortuitum
CC plasmid pAL 5000 origin of replication, ori which has been derived by PCR
CC analysis. This sequence is used in a method which results in the
CC formation of Mycobacterium recombinant vaccines from compositions that
CC provide a continuous source of protein to an animal and stimulate
CC cellular immunity. Such compositions are used to stimulate cellular
CC immunity (by inducing Th-1 cells or cytotoxic T lymphocytes),
CC particularly as vaccines (live or dead) for treating and preventing
CC diseases caused by intracellular pathogens (bacteria, viruses, rickettsia
CC or protozoa), and also cancer, autoimmune diseases, allergy and bovine
CC spongiform encephalitis, in humans or animals. The vaccines are
CC administered by injection, orally and nasally. These compositions provide
CC consistent and long-lasting immunity. Transformed cells used in the
CC method are retained within macrophages, blocking the killing mechanism
CC but producing protective immunogen which is processed and presented by
CC the macrophage. Several immunogens may be included in the same vaccine
CC and the Mycobacterium cells serve as adjuvant.
XX
SQ Sequence 1382 BP; 275 A; 464 C; 441 G; 202 T; 0 other:
AAV69313 Length: 1382 March 5, 2002 14:17 Type: N Check: 5050 ..
Found using 'seq2-3' (pappu403.key)
...
110 CGGACCATACCGGATTAATGCTGTCTACTACCAAGCGTGAGCCGCTCCCGCAGC
|-----|
170 AATTGACGAGCTGGCTGCGCTACTGCGCGCTGGACAGCGAGCATGCTCGAGGGA
|-----|
208 213 211 216
230 TCTACCGCAAGCGCGCGCGCTAGCGCCCGCGGTACATGAGCGCAACCCACAGC
|-----|
290 CGCTGCAAAACCTGCTGCTGAGAGTAGACCATTCACAGCAGCGCTCCGAGGCTCA
293 333 338 342 347
350 GCGCCCGGGGTCCCATCCGCTGCGCAACGCGATCTGGGCAATCGGCCCAAGCGCCACG
354
410 CACACGCACTGTGGGCACTCAACGCCCTGTTCACAGCAGCAGGAATACGCGCGCTAAGC
|-----|
430 435
470 CGCTCGCATATAGCGCGCGCTGGCCGCAAGGCTTGGCGCGCGCGTGAGCGGACGCA
|-----|
520 525
530 GTTACTAGGCGCTCATGACCAAAAACCGCGCACATCGCGGAGAAAGGAATGGCTCC
565 570
590 ACTAGATCTCTACACTACAGCCACATCGAGGCGGAGCTGGGGCGCAATGCCACCGC
|-----|
650 GCGCTGCGCTACAGCAGCATACAAAGCGGCTCGGACCGCGCTAGAGGGAATTCGCG
656 661 687 692
710 CACTGTTCGATTCGCTAGGTTGTGGGCTATTCGTCGCGCGCTCATGCGGATCTACCTGC
740 745
770 CGACCGGAACGTGAGGAGACTCGCGCGCGATCTATGCGGAGTGCACGCGCAAGACG
|-----|

```

826
830      -|      |-----|
      CCGAATTCCTCGTGCAGCAGTGTGTCCCGGACCGCTACCGGACGAGCGAGTCCCGGCCA
      831      844 849
890      -----|      |-----|
      TCGGCAACAGCATTTGGCGTTGATACACAAACCAAGTCCGGCATTTTGGCGGACGGGATCG
      894      905 910
950      TGTCTACGAG
...
1027      ACAATTGGCGGGCGCGCAAAAGTCCGGCTGCACCATGAGGCAATTGCTATGAGCGAGGCT
      1077 1082
1087      |-----|      |-----|
      ACAGCGAGCGCTACGAGCGGCTACAAACCGGACCGCACTGTCCGCCAANAAGCCGTGAC
      1089 1094 1101 1106
1147      |-----|      |-----|      |-----|
      GCGCGGAAGGGCTCGAATCACCGGACTATCCGAACGCCACGCTGCGCGCTGAGGCGC
      1155 1160      1180 1185      1193
1207      AGAAGCAGCAGGAGTGGCTGCGCCGACGAGGCTGCACCGCGCGGAAGCATCCGCCCTATTC
      1267      |-----|
      ACGACGACGAGGGCCACTCTTGGCCGCAAAACGCCAACAATTTGGGCTGCATCTGACACA
      1269 1274
1327      |-----|
      CCGTTAAGCGACTCGCTATTCGGCGGAGAAAGAGCGTGGCGGAGAACGAGAGGCG
      1333 1338
-----
17 matches found in sequence:
aaav69314 : Mycobacteriophage D29 atp and int genomic DNA.
      (from "mycobacteriophg.seq")
      TOIG of: aaav69314 check: 4407 from: 1 to: 1631
ID      AAV69314 standard; DNA; 1631 BP.
XX
AC      AAV69314:
XX
DT      01-FEB-1999 (first entry)
XX
DE      Mycobacteriophage D29 atp and int genomic DNA.
XX
KW      Vaccine: cellular immunity; Th-1 cell; cancer; cytotoxic T lymphocyte;
KW      treatment; prevention; disease; autoimmune; intracellular pathogen; BSE;
KW      allergy; bovine spongiform encephalitis; macrophage; immunogen; adjuvant;
XX      atp; int; integrase; attachment site; ds.
XX
OS      Mycobacteriophage D29.
XX
FH      Key      Location/Qualifiers
FT      misc_feature      214..256
FT      /*tag= a
FT      /note= "atp attachment site"
FT      -35_signal      322..327
FT      /*tag= b
FT      -10_signal      344..348
FT      /*tag= c
FT      RBS      367..371
FT      /*tag= d
FT      CDS      407..1529
FT      /*tag= e
FT      /product= "int"
FT      /note= "integrase"
XX
PN      MO9844096-AZ.

```

```

XX      PD      31-OCT-1998.
XX      XX
XX      PF      27-MAR-1998; 98WO-US06056.
XX      XX
XX      PR      28-MAR-1997; 97US-0042849.
XX      XX
XX      PA      (CYTO-) CYTOCLONAL PHARM INC.
XX      PI      Labidi AH:
XX      DR      WPI: 1998-568277/48.
XX      XX
XX      PR      Recombinant non-pathogenic Mycobacterium as vaccines providing long
XX      PR      term cellular immunity - useful against intracellular pathogens,
XX      PR      cancer and autoimmune disease, and are retained in host macrophages
XX      PS      Claim 3; Fig 4A; 120pp; English.
XX      CC      This genomic DNA sequence encodes the Mycobacteriophage D29 integrase,
XX      CC      int protein and also contains the atp attachment site. This sequence is
XX      CC      used in a method which results in the formation of Mycobacterium
XX      CC      recombinant vaccines from compositions that provide a continuous source
XX      CC      of protein to an animal and stimulate cellular immunity. Such
XX      CC      compositions are used to stimulate cellular immunity (by inducing Th-1
XX      CC      cells or cytotoxic T lymphocytes), particularly as vaccines (live or
XX      CC      dead) for treating and preventing diseases caused by intracellular
XX      CC      pathogens (bacteria, viruses, rickettsia or protozoa), and also cancer,
XX      CC      autoimmune diseases, allergy and bovine spongiform encephalitis, in
XX      CC      humans or animals. The vaccines are administered by injection, orally and
XX      CC      nasally. These compositions provide consistent and long-lasting immunity.
XX      CC      Transformed cells used in the method are retained within macrophages,
XX      CC      blocking the killing mechanism but producing protective immunogen which
XX      CC      is processed and presented by the macrophage. Several immunogens may be
XX      CC      included in the same vaccine and the Mycobacterium cells serve as
XX      CC      adjuvant.
XX      SQ      Sequence 1631 BP; 331 A; 485 C; 553 G; 262 T; 0 other:
AAV69314 Length: 1631 March 5, 2002 14:17 Type: N Check: 4407 ..
Found using 'seq2-3' (pappu403.key)
...
58      AGCAGCGGGGGGCTTCTTCTGCTGCGAGTGAAGTAGACAGATGTCTCGCTGAT
      108 113
118      CCGGGCAGCATTAATGCAAGTCATTAGTCTGCTTAAGTGCAGGCCCCCTTCGGGGA
      146 151
178      TCCGCTCCTCGGCGCTAAAAACACACCTCTGACCTGTGAGCGCGGAGCGGAATGAACCC
      219 224
238      GCGTAGCTAGTTTGGAAAGTAAGGGGTGCGCGTCTCA
...
310      CCTGTGCGACCTTGAAGTGAGACAAAATGCTGTTCACGAGACGCAAAAGACGCTTGAA
      360 365
370      GGTGCAATAAGTGTGCGATTCGCGTAGCTGTGTTTCGCATGCGACAGCAAGAGAGAGATG
      433 438
430      GGGATCGCTCGGACCGACGCGAGCGGTGAGTGCAGCGCGTGTACGTACGACCCGATCGA
      466 471
490      CGGCGACGGGTACTTTCGGGGCGGAGACTACGACACCGGATGAGCGCGAGAGCGTGGCT
      533 538

```

```

550      CGCGTCTGAGAAAGCGGCTGATGACACAACGAGAGTGGAC
...
609      AGAAGGCTCGGGCGAGTGGCCATCAGCGTCGAGAGTACACCAAGAAGTGGATGCGCGAGC
                                         |-----|
                                         659 664
669      GAGACCTCGCTGGCGGCGACCAAGATCTCTACAGCAGCAGCTCGCAAGCGGATCTACC
729      CGGTTGGGCGACACCGCGTCCGCGGATGACCCCCGCCCTTGTCCGGGCGTGGTGGG
          |-----|
          737 742 749 754
789      CCGGGATGGGTAAGCA
...
1028     TTGCGTGAAGCTGATGAGATCCGCGCGCAAGACATGCTGATGACGGCGGAGAGATGAAG
                                         |-----|
                                         1078 1083
1088     CTCGCCGTGGCGCGCGCGCGCGCGCGCTGCGGAGAGAAGATCGTGGTGGCGCAACACCAAG
                                         |-----|
                                         1127 1132
                                         1130 1135
1148     ACCGTGAGTCCAAAGCGCGGTGACCGCTGCGCGCTCAGCTGCGGCGGATGATCCGCGAG
                                         |-----|
                                         1192 1197
1208     CACATGGCTGACCGGACGAAAGATGAACAAGGGCGCGGAAG
...
1349     ATCCACGACCTCCGGGCGGTGGAGCCAGCCGCTGCGCGCTGCGGAGACCAAG
                                         |-----|
                                         1399 1404
1409     GAGCTGATGGTGGCGCTCGGGGACACGAGACTCCGCGCATGGCGATGAGTACGATGCGC
                                         |-----|
                                         1447 1452
1469     TCAGCAGCCCGTGCAGAGAGAGATAGCGAGCGGAA
...
1575     GGGGTTTCTTTCACCTCACTCACTCCACACGCTCCATTGGATCTTGGGGCTGTAGACGATC
                                         |-----|
                                         1625 1630
-----
15 matches found in sequence:
aav69315 : Mycobacteriophage D29 atpD and int genomic DNA derived by PCR #1.
(From "Mycobacteriophages")
FOIG of: aav69315 check: 6993 from: 1 to: 1413
ID      AAV69315 standard; DNA: 1413 BP.
XX
AC      AAV69315;
XX
DT      01-FEB-1999 (first entry)
XX
DE      Mycobacteriophage D29 atpD and int genomic DNA derived by PCR #1.
XX
KW      Vaccine; cellular immunity; Th-1 cell; cancer; cytotoxic T lymphocyte;
KW      treatment; prevention; disease; autoimmune; intracellular pathogen; BSE;
KW      allergy; bovine spongiform encephalitis; macrophage; immunogen; adjuvant;
KW      atp; int; integrase; attachment site; ds.
XX
OS      Mycobacteriophage D29.

```

```

XX      Key      Location/Qualifiers
FH      misc_feature      96..138
FT      /tag= a
FT      /note= "atp attachment site"
FT      -35_signal      204..209
FT      /*tag= b
FT      -10_signal      226..230
FT      /*tag= c
FT      RBS      249..253
FT      /*tag= d
FT      CDS      289..1413
FT      /*tag= e
FT      /*product= "Int"
FT      /*note= "Integrase"
XX
XX      WO9844096-A2.
XX
XX      31-OCT-1998.
XX
XX      27-MAR-1998; 98NC-US06056.
XX
XX      28-MAR-1997; 97US-0042849.
XX
XX      (CYTO-) CYTOCLONAL PHARM INC.
XX
XX      Labidi AH:
XX
XX      WPI: 1998-568277/48.
XX
XX      Recombinant non-pathogenic Mycobacterium as vaccines providing long
XX      term cellular immunity - useful against intracellular pathogens,
XX      cancer and autoimmune disease, and are retained in host macrophages
XX
XX      Claim 4; Fig 4B; 120pp: English.
XX
XX      This genomic DNA sequence encodes the Mycobacteriophage D29 integrase,
XX      int protein and also contains the atp attachment site. This sequence is
XX      used in a method which results in the formation of Mycobacterium
XX      recombinant vaccines from compositions that provide a continuous source
XX      of protein to an animal and stimulate cellular immunity. Such
XX      compositions are used to stimulate cellular immunity (by inducing Th-1
XX      cells or cytotoxic T lymphocytes), particularly as vaccines (live or
XX      dead) for treating and preventing diseases caused by intracellular
XX      pathogens (bacteria, viruses, rickettsia or protozoa), and also cancer,
XX      autoimmune diseases, allergy and bovine spongiform encephalitis, in
XX      humans or animals. The vaccines are administered by injection, orally and
XX      nasally. These compositions provide consistent and long-lasting immunity.
XX      Transformed cells used in the method are retained within macrophages,
XX      blocking the killing mechanism but producing protective immunogen which
XX      is processed and presented by the macrophage. Several immunogens may be
XX      included in the same vaccine and the Mycobacterium cells serve as
XX      adjuvant.
XX
XX      Sequence 1413 BP; 289 A; 423 C; 490 G; 211 T; 0 other;
XX
XX      AAV69315 Length: 1413 March 5, 2002 14:17 Type: N Check: 6993 ..
XX      Found using 'seq2-3' (pappu403.key)
1      CGGGCAGCATAAATGACGATTAAGTGCCTTAAGTGCAGGCGCGCCCTCTGGGGAT
          |-----|
          28 33
61      CCGGTCTCGGGCTAAACACCACTCTGACCTGTGAGCGGCGACGGGAATGAACCCG
          |-----|
          101 106
121      CGTAGCTAGTTTGAAGTAGAGGGGTGCGGCTGTCA
...
192      CCTGTGCGACCTTGAAGTGGACAAAATGCTGTTTCAGCGACGCAAGCAAGAGCTTGAA
          |-----|

```

```

252 GGTGCATTAAGTTCGATTCGGTAGCCTGTTTCGATGCGACGACAGACGAGAGATG      242 247
      |-----|
312 GGGATCGCTGCGGAGCCAGCCAGCGGTGAGTGCAGCGTGTACGTACAGCCCATCGA      348 353
      315 320
      |-----|
372 CGGGCAGCGGTACTTTCGGGCCGAGGAACTACGACACCGGATGGAGCGGAGCCTGCT      415 420
      |-----|
432 CGCGTGTGAGAGCGCGGTGATGACAAACGAGAGTGAGC
      |-----|
491 AGAAGGTGCGGCGAGTGCATCAAGGTGAGAGATACACCAAGAAAGTGCAGCGCAGC      541 546
      |-----|
551 GAGACCTCGGTGGGGGACCAAGATCTCTACAGCAGCAGCGTCCGCAAGCGGATCTACC
      |-----|
611 CGGTGTTGGGAGACCCCGTCCGAGATGACCCCGCCCTTGTCCGGGCGTGTGGG      619 624 631 636
      |-----|
671 CCGGGATGGGTAAAGCA
      |-----|
910 TTGGGTAGCTGATCGATCCGCCGCAAGACATCGTGTGATGAGCGGAGACGATGAG      960 965
      |-----|
970 CTCGCGTGGCGCGGGGCGGCGCCGCGTGCAGAGAAAGTGTGCGGACACACCAAG      1009 1014
      |-----|
      1012 1017
1030 ACCGTGAGTTCAAAGCGCGCGGTGACCGCTCCGCTCAGCTCCGCGGATGCGCGAG      1074 1079
      |-----|
1090 CACATGGCTGACCGGAGCAAGATGATCAACAGGGGCCGGAAG
      |-----|
1231 ATCCAGACCTCCGGGCGGTGGAGCCAGCTGGCGGCTCAGGCGCGTGGAGACCAAG      1281 1286
      |-----|
1291 GAGCTGATGTTGGCCTCTCGGGGACAGACTCCGCGCATGGCGATGAGTACCAAGATGCC      1329 1334
      |-----|
1351 TCAGCAGCCCGTGTACGAGGAGATAGCGAGCGGAA
      |-----|
14 matches found in sequence:
aav69316 : Mycobacteriophage D29 attP and int genomic DNA derived by PCR #2.
(from "Mycobacteriophg.seq")
TOIG of: aav69316 check: 7788 from: 1 to: 1374
ID AAV69316 standard; DNA: 1374 BP.
XX
AC AAV69316;
XX
DT 01-FEB-1999 (first entry)
XX
DE Mycobacteriophage D29 attP and int genomic DNA derived by PCR #2.

```

```

XX Vaccine; cellular immunity; Th-1 cell; cancer; cytotoxic T lymphocyte;
KW treatment; prevention; disease; autoimmune; intracellular pathogen; BSE;
KW allergy; bovine spongiform encephalitis; macrophage; immunogen; adjuvant;
KM attP; int; integrase; attachment site; ds.
XX Mycobacteriophage D29.
OS
FH Key Location/Qualifiers
FH misc_feature 57..99
FH /tag= a
FH /note= "attP attachment site"
FH -35_signal 165..170
FH /tag= b
FH -10_signal 187..191
FH /tag= c
FH RBS 210..214
FH /tag= e
FH CDS 250..1374
FH /tag= f
FH /product= "int"
FH /note= "integrase"
FH
XX W09844096-A2.
XX 31-OCT-1998.
XX 27-MAR-1998; 98WO-US06056.
XX 28-MAR-1997; 97US-0042849.
XX (CYTO-) CYTOCLONAL PHARM INC.
XX LabId: AH;
XX WPI: 1998-568277/48.
XX
XX This genomic DNA sequence encodes the Mycobacteriophage D29 integrase,
XX int protein and also contains the attP attachment site. This sequence is
XX used in a method which results in the formation of Mycobacterium
XX recombinant vaccines from compositions that provide a continuous source
XX of protein to an animal and stimulate cellular immunity. Such
XX compositions are used to stimulate cellular immunity (by inducing Th-1
XX cells or cytotoxic T lymphocytes), particularly as vaccines (live or
XX dead) for treating and preventing diseases caused by intracellular
XX pathogens (bacteria, viruses, rickettsia or protozoa), and also cancer,
XX autoimmune diseases, allergy and bovine spongiform encephalitis, in
XX humans or animals. The vaccines are administered by injection, orally and
XX nasally. These compositions provide consistent and long-lasting immunity.
XX Transformed cells used in the method are retained within macrophages,
XX blocking the killing mechanism but producing protective immunogen which
XX is processed and presented by the macrophage. Several immunogens may be
XX included in the same vaccine and the Mycobacterium cells serve as
XX adjuvant.
XX
XX Sequence 1374 BP; 279 A; 415 C; 478 G; 202 T; 0 other;
XX
AAV69316 Length: 1374 March 5, 2002 14:17 Type: N Check: 7788
Found using 'seq2-3' (pappu403.key)
12 TCTCGGGGATCGGTCTCGGGCTAAACACACTCTGACCTGTGAGCGGGGAGCGGA      62 67
72 ATGAACCCGCGTAGCTAGTTTGGAAGTAAGGGGTGCGCGCTGCA

```

```

...
153 CCCTGTGCGACCTTGAAGTGGACAAAAATCCCTGTTTCAGGGACACGCCAAGACGCTTGAA
      |-----|
213 GGTGCAATPAAAGTCCGATTCGGTAGCCTGTTTCGATGGCAGCAAGACGAGAGAGATG
      |-----|
273 GGGATGCTGCGGAGCCAGCCGACGGCTGAGTGCAGCGTGTACGTCAGCCGATCGA
      276 281 |-----|
      309 314 |-----|
333 CGGGCAGCGGTACTTGGGGCCGAGAACTACGACAAACCGGATGACCGCGAGCTGGCT
      |-----|
393 CGCGCTGAGAAAGCGGCTGATCGACAACGAGAGTGGAC
      |-----|
...
452 AGAAGGCTGGCGGAGTGCATCAGGTCGAGGAGTACACCAAGAAAGTGGATGCGCGAGC
      |-----|
512 GAGACCTCGGCTGGCGGACCAAGGATCTTACAGCAACGACGCTCGCAACGGGATCTTACC
      |-----|
572 CGGTGTTGGGCGACACCCCGGTCGCGAGATGACCCCGCCCTTGTCCGGGGCGTGTGGG
      580 585 |-----|
      592 597 |-----|
632 CCGGATGGGTAAGCA
...
871 TTGCGTAGCTGATCGAGATCCGCCGCAAGACATGCTGGATGACGGCGAGAGATGAAG
      |-----|
931 CTCGGGTCGCGCGGGCGGGCCGGCGTGGCGGAGAAAGATCGTCTGCGGCAACACCAAG
      970 975 |-----|
      973 978 |-----|
991 ACCGTCAGGTCCAAGGGCCGGTGAACCCGCTCAGCTCGCGGGGATGATCCGCGAG
      |-----|
1051 CACATGGCTGACCGGAGCAAGATGAACAAGGGCCGGAAG
      |-----|
...
1192 ATCCACGACCTCCGGGCGCTGGAGAGCCAGCTGGCGGCTCAGCGCGGTCGACGACCAAG
      |-----|
1252 GAGCTGATGTTGGCCTTCGGGCAACAGACTCCGGGATGGAGTGAAGTACCAAGTGGCC
      1290 1295 |-----|
1312 TCAGCAGCCCGTAGCAGAGAGATAGCAGGCGAA
...
48 matches found in sequence:
aav69318 : M. fortuitum plasmid pAL 5000 genomic DNA fragment.
(from "mycobacterieng.seq")
TOIG of: aav69318 check: 8209 from: 1 to: 2096
ID AAV69318 standard; DNA: 2096 BP.
XX

```

```

AC AAV69318;
XX
XX 01-FEB-1999 (first entry)
DE M. fortuitum plasmid pAL 5000 genomic DNA fragment.
XX
XX Vaccine; cellular immunity; Th-1 cell; cancer; cytotoxic T lymphocyte;
XX treatment; prevention; disease; autoimmune; intracellular pathogen; BSE;
XX allergy; bovine spongiform encephalitis; macrophage; immunogen;
XX adjuvant; ds.
XX Mycobacterium fortuitum.
OS WO9844096-A2.
XX
XX
XX 31-OCT-1998.
XX
XX 27-MAR-1998; 98MO-US06056.
XX
XX 28-MAR-1997; 97US-0042849.
XX
XX (CYTO-) CYTOCLONAL PHARM INC.
XX
XX Labidi AH;
XX WPI; 1998-568277/48.
XX
XX Recombinant non-pathogenic Mycobacterium as vaccines providing long
XX term cellular immunity - useful against intracellular pathogens,
XX cancer and autoimmune disease, and are retained in host macrophages
XX
XX Disclosure; Fig 6; 120pp; English.
XX
XX This genomic DNA sequence is a fragment of the Mycobacterium fortuitum
XX plasmid pAL 5000 which contains ORF2. This sequence is used in a
XX method which results in the formation of Mycobacterium recombinant
XX vaccines from compositions that provide a continuous source of protein to
XX an animal and stimulate cellular immunity. Such compositions are used to
XX stimulate cellular immunity (by inducing Th-1 cells or cytotoxic T
XX lymphocytes), particularly as vaccines (live or dead) for treating and
XX preventing diseases caused by intracellular pathogens (bacteria, viruses,
XX rickettsia or protozoa), and also cancer, autoimmune diseases, allergy
XX and bovine spongiform encephalitis, in humans or animals. The vaccines
XX are administered by injection, orally and nasally. These compositions
XX provide consistent and long-lasting immunity. Transformed cells used in
XX the method are retained within macrophages, blocking the killing
XX mechanism but producing protective immunogen which is processed and
XX presented by the macrophage. Several immunogens may be included in the
XX same vaccine and the Mycobacterium cells serve as adjuvant.
XX
XX Sequence 2096 BP; 348 A; 743 C; 670 G; 335 T; 0 other;
XX
AAV69318 Length: 2096 March 5, 2002 14:17 Type: N Check: 8209 ..
Found using 'seq2-3' (pappu403.key)
...
15 AACCCGAGCTGCAAGTAGCGAAAGTCTCATACACCAAGAGCGGCTTGGCGATACCA
      |-----|
75 GCGCGGGGGCGAATCCCGCCAGGAATGCCGTCAATCGGTGTCCGGGACTGGCGGAG
      79 83 |-----|
      88 |-----|
135 CGGACACTCGACCAACAACAACAACGATGATAGGAGCAGCAGCAGCAGATGGA
      161 166 |-----|
      164 169 |-----|
      172 177 |-----|
      180
195 TGATCCGAATCACTGCGCTGTCCATACAGCGCGCCCTCGAAGTACACAGCTTCAT

```

```

255      GCGCCTGTGCAAGACTGCATGAGCGCTACTGCGGGCGCGTCTCAACGCACTGCTCGAA
      |-----|
      279 284

315      GAATTCGACAGCGCCAGTGCACCGAACTCCTTGTGTGTCGGCTTCAGCTCGGCGCT
      |-----|
      369 374

375      CCAGCTTTTCACCTTGGGCGCGGACAAATTCGACGACCTTGTAGCGATCGACGATTTGTT
      |-----|
      405 410 417 422 433

435      GCGCGCAGCAATGCCCCGACATCCAGTCCCTTGATCGAGTGGCGCGGCACACAG
      438

495      CTCCGCGATCCGACCCCGATCCAGCGCTGCCTCACCACTTTTCGTCGCGGGGGCTCA
      |-----|
      517 522 539 544

555      CCCGGTACTGACCGGATGCGCCACTATGAAACGGCTACGCGCGCGGACAGCGCGGCG
      |-----|
      572 577 611

615      CTGGGGGGGCGACGTTTCATCACACCGGACCGGGACACGAGCTTATCATGATGCGCG
      616 653 658

675      GCTGAATCGCGCGGCTGCTGCGGACAGCAGTCCGCGAGCTCGTCCGATCGATGTACT
      |-----|
      687 692

735      GCCGGCGCGGATGCTGTCACGAGAAATGTGGACACAGCGCTTGTGCGGGGCTC
      |-----|
      747 752 775 780

795      TTCGCGGATGAGATCCGCTCGAGGCGGCTCGCGCGGCTGTGGCATGTGGGGCG
      |-----|
      867 872

855      TGCGGCTCCCGCGCGCCCATCGGCGCCCATTTGGCATTCGGCCATGCGCGCCAT
      |-----|

915      CATTCTGTGAGCGCAGCTGGCGCGTCTTCAATGAGGACAGCCCGCTGACGCGACGT
      |-----|
      967 972

975      GAGGCGGTGCGCGCCGAATCTGGGCGCGATCACTCGGCGACCGGTGCGGATTT
      |-----|

1035      GGGCGCGGTGTGTGCGGTGCGACACACGCGCGCGCCCGCCGATCGCGCA
      |-----|
      1071 1076 1087 1092

1095      ACCACGGGGTGTGTGGTGTGCGTCCGACCTGCCAGAAATGTCCTCCGCGTCTGCT
      |-----|
      1129 1134 1142

1155      GAACACCGCGAGCGCGCCGCACTCGGAAAGCGCTGGCATATGATGTCGCGCGGGGG
      |-----|
      1186 1191

1215      CCATTCGAGCGGGTGCACCTGTGCGGGGCTGTGGCGCGCTGCGCTGCGACCA
      |-----|
      1250 1255 1262 1267 1256 1261

1275      TCGCGCGTGGGCGCGACACCTTCGATGTGCAACGCGCGCGCGCGGATGTCGTC
      |-----|
      1320 1325 1334

```

```

1335      ACGCGCGCGCGCGCGAGATGCTGCTCCCAACCGTCGCGGCTGAGGTCGCGGGCGA
      |-----|
      1335 1340 1347 1352 1361 1390

1395      CCGGAATAATGCTTATGCTGCGCGGACACCTTGAATCGGTGTCGCGCTCGTGGGCGAG
      |-----|
      1395

1455      GCTTCGCGTGCAGCTGACGACGCGCGCTCCACATGCTCCACGCGCGCTGACCTCGGG
      |-----|
      1462 1467 1487 1492 1500

1515      ATCGGACGCGTCCCGCGCGGATGACCGGGCGCGCGCGCGCTCTCCACCGCACG
      |-----|

1575      CAGGCGCGTGGCGGATTTTCAGACAGTGTGCGCGCGCCCATTTCCGACATCTTTCTCGGGG
      |-----|
      1585 1590 1600 1605

1635      CCGATCGCGCGACCGGACCGCAATGTGCTCCGGAAGAGGCTGCGCGCGATCGACTCCAGC
      |-----|
      1638 1643 1658 1663

1695      AACGCGCGCATGTGATGCTGCTCCTGAAACTCGGCGCTGTGTCAGCGGAATCGCGTCA
      |-----|
      1745 1750

1755      TAACGGAATGGGCGCGCGCGCGCGCGGATATCCAGCGCGGACGCTTTCGAACTTGGTTC
      |-----|
      1763 1768 1796 1801

1815      ATGTGTAATCCCTCGCTTGTGAACCTTGATTAAGCT
      |-----|

...

1970      AACCAAGCCCGGGAAGCGCTGCGACAGCGTAAGGCGCGCTCATGCGGCGCGCTGCGC
      |-----|
      2020 2025

2030      ACGATGTCCTGCACTTGAGCCAAAGCTCGAACAACCTGCTGTGTG
      |-----|

...

-----
1 match found in sequence:
aav80096 ; Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacteriophage")
TOIG of: aav80096 check: 8058 from: 1 to: 22

ID      AAV80096 standard; DNA; 22 BP.
XX
AC      AAV80096;
XX
DE      12-MAR-1999 (first entry)
XX
XX      Immunomodulatory oligo comprising an ISS sequence.
XX
KW      Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW      ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW      human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW      B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS      Synthetic.
XX
PN      WO9855495-A2.
XX
PD      10-DEC-1998.
XX
PE      05-JUN-1998; 98WO-US11578.
XX
PR      06-JUN-1997; 97US-0048793.
XX
PA      (DYNA-) DYNAVAX TECHNOLOGIES CORP.

```


XX Dina D, Roman M, Schwartz D;
PI
XX WPI: 1999-059898/05.
DR
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 7; Page 29; 63pp; English.
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTG,
CC GACGTTCC, and GACGTTG. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 4 C; 7 G; 5 T; 0 other;
AAV80096 Length: 22 March 5, 2002 14:19 Type: N Check: 8058 ..
Found using 'seq2-3' (pappu403.key)
1
TGACCTGTGACGTTGCGAGATGA
9 14

1 match found in sequence:
aav80097 ; Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterieng.seq")
TOIG of: aav80097 check: 8143 from: 1 to: 22
ID AAV80097 standard; DNA: 22 BP.
XX
AC AAV80097;
XX
XX 12-MAR-1999 (first entry)
DT
XX
XX Immunomodulatory oligo comprising an ISS sequence.
DE
XX
XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAXX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI: 1999-059898/05.
XX
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases

XX Claim 5; Page 29; 63pp; English.
PS
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTG,
CC GACGTTCC, and GACGTTG. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
AAV80097 Length: 22 March 5, 2002 14:19 Type: N Check: 8143 ..
Found using 'seq2-3' (pappu403.key)
1
TGACCTGTGACGTTGCGAGATGA
9 14

1 match found in sequence:
aav80098 ; Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterieng.seq")
TOIG of: aav80098 check: 9493 from: 1 to: 23
ID AAV80098 standard; DNA: 23 BP.
XX
AC AAV80098;
XX
XX 12-MAR-1999 (first entry)
DT
XX
XX Immunomodulatory oligo comprising an ISS sequence.
DE
XX
XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAXX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI: 1999-059898/05.
XX
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 6; Page 29; 63pp; English.
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTG,
CC GACGTTCC, and GACGTTG. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,

CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 23 BP; 6 A; 8 C; 3 G; 6 T; 0 other;
AAV80098 Length: 23 March 5, 2002 14:19 Type: N Check: 9493 ..
Found using 'seq2-3' (pappu403.key)
1 |-----|
TCATCTCGACGCTTCACAGTCA
9 14
1 matches found in sequence:
aav80099 ; Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterng.seq")
TOIG of: aav80099 check: 8079 from: 1 to: 22
ID AAV80099 standard; DNA; 22 BP.
XX
AC AAV80099;
XX
DT 12-MAR-1999 (first entry)
XX
DE Immunomodulatory oligo comprising an ISS sequence.
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
PN W09855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98W0-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI: 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 8; Page 29; 63pp: English.
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGG,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent

CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 4 C; 6 G; 6 T; 0 other;
AAV80099 Length: 22 March 5, 2002 14:19 Type: N Check: 8079 ..
Found using 'seq2-3' (pappu403.key)
1 |-----|
TGACTGTGACGCTTCAGATGA
9 14
2 matches found in sequence:
aav80100 ; Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterng.seq")
TOIG of: aav80100 check: 5564 from: 1 to: 26
ID AAV80100 standard; DNA; 26 BP.
XX
AC AAV80100;
XX
DT 12-MAR-1999 (first entry)
XX
DE Immunomodulatory oligo comprising an ISS sequence.
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
PN W09855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98W0-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI: 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 9; Page 29; 63pp: English.
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGG,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
XX
SQ Sequence 26 BP; 5 A; 9 C; 4 G; 8 T; 0 other;
AAV80100 Length: 26 March 5, 2002 14:19 Type: N Check: 5564 ..
Found using 'seq2-3' (pappu403.key)
1 |-----| |-----|

```
1
-----
TCCATACGTTCCGCTACGTCGTC
6      11      17      22

1 match found in sequence:
aav80101 : Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterieng.seq")
TOIG of: aav80101 check: 8079 from: 1 to: 22

ID AAV80101 standard; DNA: 22 BP.
XX
AC AAV80101;
XX
DT 12-MAR-1999 (first entry)
XX
DE Immunomodulatory oligo comprising an ISS sequence.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT modified_base 11
FT /*tag= a
FT /note= "5-bromocytosine"
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI; 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 22: Page 30; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 4 C; 6 G; 6 T; 0 other;
AAV80101 Length: 22 March 5, 2002 14:19 Type: N Check: 8079 ..
Found using 'seq2-3' (pappu403.key)
```

```
1
-----
1 match found in sequence:
aav80102 : Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterieng.seq")
TOIG of: aav80102 check: 8143 from: 1 to: 22

ID AAV80102 standard; DNA: 22 BP.
XX
AC AAV80102;
XX
DT 12-MAR-1999 (first entry)
XX
DE Immunomodulatory oligo comprising an ISS sequence.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT modified_base 11
FT /*tag= a
FT /note= "5-bromocytosine"
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI; 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 23: Page 30; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
AAV80102 Length: 22 March 5, 2002 14:19 Type: N Check: 8143 ..
Found using 'seq2-3' (pappu403.key)
```

```
1
-----
1 match found in sequence:
aav80103 : Immunomodulatory oligo comprising an ISS sequence.
(from "mycobacterieng.seq")
TOIG of: aav80103 check: 8143 from: 1 to: 22

ID AAV80103 standard; DNA: 22 BP.
XX
AC AAV80103;
XX
DT 12-MAR-1999 (first entry)
XX
DE Immunomodulatory oligo comprising an ISS sequence.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT modified_base 11
FT /*tag= a
FT /note= "5-bromocytosine"
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI; 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 24: Page 30; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
AAV80103 Length: 22 March 5, 2002 14:19 Type: N Check: 8143 ..
Found using 'seq2-3' (pappu403.key)
```

```
(from "mycobacterng.seq")
TOIG of: aav80103 check: 8143 from: 1 to: 22
ID AAV80103 standard; DNA: 22 BP.
XX
AC AAV80103;
XX
DT 12-MAR-1999 (first entry)
XX
DE Immunomodulatory oligo comprising an ISS sequence.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS; cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; trypanosoma; schistosoma.
XX
OS Synthetic.
XX
FH Key location/qualifiers
FT modified_base 11
FT /*tag= a
FT /note= "5-bromocytosine"
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI: 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Claim 24; Page 30; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTCC,
CC GACGTTCC, and GACGTTCC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, leishmania, trypanosoma and
CC schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80096 to AAV80103 represent
CC specific claimed examples of such immunomodulatory oligonucleotides.
XX
SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
AAV80103 Length: 22 March 5, 2002 14:19 Type: N Check: 8143 ..
Found using 'seq2-3' (pappu403.key)
1 TGACTGTGAACGTTGACGATGA
9 14
-----
1 match found in sequence:
aav80105 : Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterng.seq")
TOIG of: aav80105 check: 8045 from: 1 to: 22
```

```
ID AAV80105 standard; DNA: 22 BP.
XX
AC AAV80105;
XX
DT 12-MAR-1999 (first entry)
XX
DE Oligo used in experiments for stimulation of cytokine production.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS; cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; trypanosoma; schistosoma.
XX
OS Synthetic.
XX
PN WO9855495-A2.
XX
PD 10-DEC-1998.
XX
PF 05-JUN-1998; 98WO-US11578.
XX
PR 06-JUN-1997; 97US-0048793.
XX
PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI Dina D, Roman M, Schwartz D;
XX
DR WPI: 1999-059898/05.
XX
PT Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Example 1; Page 29; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTCC,
CC GACGTTCC, and GACGTTCC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, leishmania, trypanosoma and
CC schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
SQ Sequence 22 BP; 5 A; 7 C; 4 G; 6 T; 0 other;
AAV80105 Length: 22 March 5, 2002 14:19 Type: N Check: 8045 ..
Found using 'seq2-3' (pappu403.key)
1 TCATCTGGAACGTTGACGATGA
9 14
-----
2 matches found in sequence:
aav80106 : Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterng.seq")
TOIG of: aav80106 check: 8147 from: 1 to: 22
ID AAV80106 standard; DNA: 22 BP.
XX
AC AAV80106;
XX
DT 12-MAR-1999 (first entry)
```

XX Oligo used in experiments for stimulation of cytokine production.
DE Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
XX ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KM human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX WO9855495-A2.
XX PN 10-DEC-1998.
XX PD 05-JUN-1998; 98WO-US11578.
XX PF 06-JUN-1997; 97US-0048793.
XX PR (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX PA Dina D, Roman M, Schwartz D;
XX PI WPI; 1999-059898/05.
XX DR Immunostimulatory oligonucleotides regulate the immune system - and
XX PT contain an immune-stimulating octanucleotide sequence; for treating
XX cancer, allergic and infectious diseases
XX
XX Example 1; Page 29; 63pp; English.
XX
PS The invention relates to immunomodulatory oligonucleotides that comprise
XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
AAV80106 Length: 22 March 5, 2002 14:19 Type: N Check: 8147 ..
Found using 'seq2-3' (pappu403.key)

1 TGACTGTGACGTTAGCGATGA
9 14 20
15

1 match found in sequence:
aav80107; Oligo used in experiments for stimulation of cytokine production.
(from "mycobactering.seq")
TolG of: aav80107 check: 8155 from: 1 to: 22

ID AAV80107 standard; DNA: 22 BP.
XX
XX AAV80107;
XX
XX 12-MAR-1999 (first entry)
XX
XX Oligo used in experiments for stimulation of cytokine production.
XX

KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KM ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX WO9855495-A2.
XX PN 10-DEC-1998.
XX PD 05-JUN-1998; 98WO-US11578.
XX PF 06-JUN-1997; 97US-0048793.
XX PR (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX PA Dina D, Roman M, Schwartz D;
XX PI WPI; 1999-059898/05.
XX DR Immunostimulatory oligonucleotides regulate the immune system - and
XX PT contain an immune-stimulating octanucleotide sequence; for treating
XX cancer, allergic and infectious diseases
XX
XX Example 1; Page 29; 63pp; English.
XX
PS The invention relates to immunomodulatory oligonucleotides that comprise
XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
SQ Sequence 22 BP; 6 A; 3 C; 7 G; 6 T; 0 other;
AAV80107 Length: 22 March 5, 2002 14:19 Type: N Check: 8155 ..
Found using 'seq2-3' (pappu403.key)

1 TGACTGTGACGTTAGCGATGA
9 14 21
15

1 match found in sequence:
aav80108; Oligo used in experiments for stimulation of cytokine production.
(from "mycobactering.seq")
TolG of: aav80108 check: 6474 from: 1 to: 21

ID AAV80108 standard; DNA: 21 BP.
XX
XX AAV80108;
XX
XX 12-MAR-1999 (first entry)
XX
XX Oligo used in experiments for stimulation of cytokine production.
XX
XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX

```
OS Synthetic.
XX
XX PN WO9855495-A2.
XX
XX PD 10-DEC-1998.
XX
XX PF 05-JUN-1998; 98WO-US11578.
XX
XX PR 06-JUN-1997; 97US-0048793.
XX
XX PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
XX PI Dina D, Roman M, Schwartz D;
XX
XX DR WPI: 1999-059898/05.
XX
XX PT Immunostimulatory oligonucleotides regulate the immune system - and
XX contain an immune-stimulating octanucleotide sequence; for treating
XX cancer, allergic and infectious diseases
XX
XX PS Example 1; Page 29; 63pp; English.
XX
XX CC The invention relates to immunomodulatory oligonucleotides that comprise
XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
XX sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
XX GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
XX patients needing immune regulation, such as those suffering from cancer,
XX an allergic disease and asthma. They are also used to prevent infectious
XX diseases such as influenza, herpes, hepatitis B, human immunodeficiency
XX and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
XX Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
XX Schistosoma. The immunomodulatory sequences are used to screen for human
XX CC oligonucleotide; and determining the relative amount of Th1-biased
XX CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
XX CC oligonucleotides that were tested for immunostimulatory activity. These
XX CC were used in experiments for the stimulation of cytokine production and
XX CC were found to lack immunostimulatory activity. The invention provides
XX CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
XX SQ Sequence 21 BP; 7 A; 2 C; 7 G; 5 T; 0 other;
AAV80108 Length: 21 March 5, 2002 14:19 Type: N Check: 6474 ..
Found using 'seq2-3' (pappu403.key)

1 TGACGTGACGTTAGAGATGA
8 13
-----
1 match found in sequence:
aav80109; Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterin.seq")
TOIG of: aav80109 check: 9734 from: 1 to: 23

ID AAV80109 standard; DNA: 23 BP.
XX
XX AC AAV80109;
XX
XX DT 12-MAR-1999 (first entry)
XX
XX DE Oligo used in experiments for stimulation of cytokine production.
XX
XX KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
XX ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
XX human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
XX B. pertussis; malaria; plasmodia; leishmania; trypanosoma; schistosoma.
XX
XX OS Synthetic.
XX
XX FT WO9855495-A2.
XX
XX PN 10-DEC-1998.
XX
XX PD
```

```
XX
XX PF 05-JUN-1998; 98WO-US11578.
XX
XX XX
XX PR 06-JUN-1997; 97US-0048793.
XX
XX PA (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
XX PI Dina D, Roman M, Schwartz D;
XX
XX DR WPI: 1999-059898/05.
XX
XX PT Immunostimulatory oligonucleotides regulate the immune system - and
XX contain an immune-stimulating octanucleotide sequence; for treating
XX cancer, allergic and infectious diseases
XX
XX PS Example 1; Page 29; 63pp; English.
XX
XX CC The invention relates to immunomodulatory oligonucleotides that comprise
XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
XX sequences are selected from the group consisting of AACGTTCC, AACGTTCCG,
XX GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
XX patients needing immune regulation, such as those suffering from cancer,
XX an allergic disease and asthma. They are also used to prevent infectious
XX diseases such as influenza, herpes, hepatitis B, human immunodeficiency
XX and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
XX Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
XX Schistosoma. The immunomodulatory sequences are used to screen for human
XX CC oligonucleotide; and determining the relative amount of Th1-biased
XX CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
XX CC oligonucleotides that were tested for immunostimulatory activity. These
XX CC were used in experiments for the stimulation of cytokine production and
XX CC were found to lack immunostimulatory activity. The invention provides
XX CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
XX SQ Sequence 23 BP; 7 A; 3 C; 7 G; 6 T; 0 other;
AAV80109 Length: 23 March 5, 2002 14:19 Type: N Check: 9734 ..
Found using 'seq2-3' (pappu403.key)

1 TGACGTGACGTTAGAGATGA
10 15
-----
1 match found in sequence:
aav80112; Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterin.seq")
TOIG of: aav80112 check: 5590 from: 1 to: 20

ID AAV80112 standard; DNA: 20 BP.
XX
XX AC AAV80112;
XX
XX DT 12-MAR-1999 (first entry)
XX
XX DE Oligo used in experiments for stimulation of cytokine production.
XX
XX KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
XX ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
XX human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
XX B. pertussis; malaria; plasmodia; leishmania; trypanosoma; schistosoma.
XX
XX OS Synthetic.
XX
XX FH Key location/Qualifiers
XX FT modified_base 8 /*tag= a
XX FT /note= "5-bromocytosine"
XX
XX PN WO9855495-A2.
XX
XX PD 10-DEC-1998.
XX
```

```
XX 05-JUN-1998; 98MO-US11578.
PF
XX 06-JUN-1997; 97US-0048793.
PR
XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
PA
XX Dina D, Roman M, Schwartz D;
PI
XX WPI; 1999-059898/05.
DR
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Example 2; Page 30; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTC, AACGTTCG,
CC GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
SQ Sequence 20 BP; 3 A; 5 G; 5 G; 7 T; 0 other;
AAV80112 Length: 20 March 5, 2002 14:19 Type: N Check: 5590 ..
Found using 'seq2-3' (pappu403.key)
1 |-----|
TCCATGACGTTCTGATGCT
6 11
1 match found in sequence:
aav80113; Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterieng.seq")
TOIG of: aav80113 check: 5498 from: 1 to: 20
ID AAV80113 standard; DNA: 20 BP.
XX
AC AAV80113:
XX
XX 12-MAR-1999 (first entry)
DT
XX
DE Oligo used in experiments for stimulation of cytokine production.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH modified_base 8
FT /*tag= a
FT /note= "5-bromocytosine"
XX
XX WO9855495-A2.
XX
XX 10-DEC-1998.
XX
```

```
XX 05-JUN-1998; 98MO-US11578.
PF
XX 06-JUN-1997; 97US-0048793.
PR
XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
PA
XX Dina D, Roman M, Schwartz D;
PI
XX WPI; 1999-059898/05.
DR
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
PS Example 2; Page 30; 63pp; English.
XX
CC The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTC, AACGTTCG,
CC GACGTTCC, and GACGTTCCG. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
SQ Sequence 20 BP; 4 A; 6 C; 3 G; 7 T; 0 other;
AAV80113 Length: 20 March 5, 2002 14:19 Type: N Check: 5498 ..
Found using 'seq2-3' (pappu403.key)
1 |-----|
TCCATGACGTTCTGATGCT
6 11
1 match found in sequence:
aav80114; Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterieng.seq")
TOIG of: aav80114 check: 5550 from: 1 to: 20
ID AAV80114 standard; DNA: 20 BP.
XX
AC AAV80114:
XX
XX 12-MAR-1999 (first entry)
DT
XX
DE Oligo used in experiments for stimulation of cytokine production.
XX
KW Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; leishmania; Trypanosoma; Schistosoma.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH modified_base 8
FT /*tag= a
FT /note= "5-bromocytosine"
XX
XX WO9855495-A2.
XX
XX 10-DEC-1998.
XX
```

```
XX 05-JUN-1998; 98MO-US11578.
PF
XX
XX 06-JUN-1997; 97US-0048793.
PR
XX
XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
PA
XX
XX Dina D, Roman M, Schwartz D;
PI
XX
XX WPI: 1999-059898/05.
DR
XX
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
XX
XX Example 2; Page 30; 63pp; English.
PS
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
XX
XX Sequence 20 BP; 4 A; 5 C; 4 G; 7 T; 0 other;
SQ
AAV80114 Length: 20 March 5, 2002 14:19 Type: N Check: 5550
Found using 'seq2-3' (pappu403.key)
1
1 |-----|
TCCATACGTTGCGATGCT
6 11
2 matches found in sequence:
aav80115; Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterng.seq")
TOIG of: aav80115 check: 1722 from: 1 to: 24
ID AAV80115 standard; DNA; 24 BP.
XX
XX AAV80115;
AC
XX
XX 12-MAR-1999 (first entry)
DT
XX
XX Oligo used in experiments for stimulation of cytokine production.
DE
XX
XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH modified_base 8 /*tag= a
FT /note= "5-bromocytosine"
FT
XX
XX WO9855495-A2.
XX
XX 10-DEC-1998.
PD
```

```
XX 05-JUN-1998; 98MO-US11578.
PF
XX
XX 06-JUN-1997; 97US-0048793.
PR
XX
XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
PA
XX
XX Dina D, Roman M, Schwartz D;
PI
XX
XX WPI: 1999-059898/05.
DR
XX
XX Immunostimulatory oligonucleotides regulate the immune system - and
PT contain an immune-stimulating octanucleotide sequence; for treating
PT cancer, allergic and infectious diseases
XX
XX
XX Example 2; Page 30; 63pp; English.
PS
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
CC at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
CC sequences are selected from the group consisting of AACGTTCC, AACGTTGC,
CC GACGTTCC, and GACGTTGC. The immunomodulatory sequences are used to treat
CC patients needing immune regulation, such as those suffering from cancer,
CC an allergic disease and asthma. They are also used to prevent infectious
CC diseases such as influenza, herpes, hepatitis B, human immunodeficiency
CC and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
CC Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
CC Schistosoma. The immunomodulatory sequences are used to screen for human
CC immunostimulatory activity by incubating macrophage cells and the
CC oligonucleotide; and determining the relative amount of Th1-biased
CC cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
CC oligonucleotides that were tested for immunostimulatory activity. These
CC were used in experiments for the stimulation of cytokine production and
CC were found to lack immunostimulatory activity. The invention provides
CC specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
XX
XX Sequence 24 BP; 5 A; 8 C; 4 G; 7 T; 0 other;
SQ
AAV80115 Length: 24 March 5, 2002 14:19 Type: N Check: 1722
Found using 'seq2-3' (pappu403.key)
1
1 |-----| |-----|
TCCATACGTTGCGCTACGTTGC
6 11 17 22
2 matches found in sequence:
aav80116; Oligo used in experiments for stimulation of cytokine production.
(from "mycobacterng.seq")
TOIG of: aav80116 check: 1722 from: 1 to: 24
ID AAV80116 standard; DNA; 24 BP.
XX
XX AAV80116;
AC
XX
XX 12-MAR-1999 (first entry)
DT
XX
XX Oligo used in experiments for stimulation of cytokine production.
DE
XX
XX Immunomodulatory; immunostimulatory; octanucleotide; immune regulation;
KW ISS: cancer; allergy; asthma; hepatitis B infection; papillomavirus;
KW human immunodeficiency virus; influenza; herpes; M. tuberculosis; ss;
KW B. pertussis; malaria; plasmodia; Leishmania; Trypanosoma; Schistosoma.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH modified_base 8 /*tag= a
FT /note= "5-bromocytosine"
FT
XX
XX modified_base 19 /*tag= b
FT /note= "5-bromocytosine"
FT
XX
XX
```



```
PN WO855495-A2.
XX
XX 10-DEC-1998.
XX
XX 05-JUN-1998; 98WO-US11578.
XX
XX 06-JUN-1997; 97US-0048793.
XX
XX (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
XX Dina D, Roman M, Schwartz D;
XX
XX WPI; 1999-059898/05.
XX
XX Immunostimulatory oligonucleotides regulate the immune system - and
XX contain an immune-stimulating octanucleotide sequence; for treating
XX cancer, allergic and infectious diseases
XX
XX Example 2; Page 30; 63pp; English.
XX
XX The invention relates to immunomodulatory oligonucleotides that comprise
XX at least 1 immunostimulatory octanucleotide sequence (ISS) where the ISS
XX sequences are selected from the group consisting of AACGTTC, AACGTTCG,
XX GACGTTC, and GACGTTCG. The immunomodulatory sequences are used to treat
XX patients needing immune regulation, such as those suffering from cancer,
XX an allergic disease and asthma. They are also used to prevent infectious
XX diseases such as influenza, herpes, hepatitis B, human immunodeficiency
XX and papillomavirus, Hemophilus influenza, Mycobacterium tuberculosis and
XX Bordetella pertussis, malarial plasmodia, Leishmania, Trypanosoma and
XX Schistosoma. The immunomodulatory sequences are used to screen for human
XX immunostimulatory activity by incubating macrophage cells and the
XX oligonucleotide; and determining the relative amount of Th1-biased
XX cytokines in the supernatant. Sequences AAV80104 to AAV80116 represent
XX oligonucleotides that were tested for immunostimulatory activity. These
XX were used in experiments for the stimulation of cytokine production and
XX were found to lack immunostimulatory activity. The invention provides
XX specific claimed examples (AAV80096-103) of immunomodulatory sequences.
XX
XX Sequence 24 BP; 5 A; 8 C; 4 G; 7 T; 0 other;
XX
AAV80116 Length: 24 March 5, 2002 14:19 Type: N Check: 1722 ..
Found using 'seq2-3' (pappu403.key)
1
1 TCCATACGTTCCGCTAACGTTG
6 11 17 22
-----
52 matches found in sequence:
aav83182: Intermediate retroviral vector pLL derived from pLNL6.
(from "mycobacterieng.seq")
TOIG of: aav83182 check: 1929 from: 1 to: 6145
ID AAV83182 standard; DNA: 6145 BP.
XX
XX AAV83182:
AC
XX 02-MAR-1999 (first entry)
DE
XX Intermediate retroviral vector pLL derived from pLNL6.
XX
XX Mouse; immunodeficient; pathogen: tumour; lymphocyte; antigen;
XX immunomodulatory; vector; vaccine; cancer; HIV; Leishmania;
XX Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
XX human immunodeficiency virus; ds.
XX
XX Synthetic.
XX
XX WO9844788-A2.
XX
XX 15-OCT-1998.
XX
XX 09-APR-1998; 98WO-US06944.
```

```
XX
XX 09-DEC-1997; 97US-0069163.
XX
XX 09-APR-1997; 97US-0838702.
XX
XX 01-MAY-1997; 97US-0848760.
XX
XX (CHAN/) CHANG L.
XX
XX Chang L;
XX
XX WPI; 1999-024005/02.
XX
XX Use of immunodeficient mice comprising human cells - particularly
XX SCID/beige mice comprising human immune cells for evaluating
XX vaccines against cancers or human pathogens, e.g. HIV
XX
XX Example 1; Page 98-101; 154pp; English.
XX
XX Immunodeficient mice comprising human cells can be used for
XX exposure to human pathogens and/or their components or human
XX tumour cells and human peripheral blood lymphocytes. Also claimed
XX is a vaccine comprising a cell modified to express an antigen and
XX an immune-modulating protein, this is preferably an expression
XX vector comprising a polynucleotide sequence that encodes the
XX antigen and immune-modulating protein. Such vectors can be used
XX to treat a subject having a tumour by transferring the expression
XX vector into the tumour so that the antigen and the immune-modulator
XX are expressed by at least the tumour. The methods can be used for
XX producing and evaluating vaccines including cancer vaccines and
XX vaccines directed against human pathogens, e.g. HIV, Leishmania,
XX Mycobacterium, Listeria or Plasmodium. This sequence is an
XX intermediate retroviral vector derived from pLNL6, a vector approved
XX for clinical use in the United States. pLL is essentially pLNL6
XX digested with ClaI and BclI to remove cloning sites and the pLNL6
XX internal Sneo gene. These were replaced with a polylinker. A neo
XX gene was then inserted under the transcriptional control of the SV40
XX enhancer/promoter to create the vector pLSN.
XX
XX Sequence 6145 BP; 1397 A; 1700 C; 1607 G; 1441 T; 0 other;
XX
AAV83182 Length: 6145 March 5, 2002 14:17 Type: N Check: 1929 ..
Found using 'seq2-3' (pappu403.key)
...
135 TTTCTTTGCTTTGAAAGACCCCGGAGTGGAAGCTAGCTTAAGTACCCGACTT
185 190
195 TGCAAGGCATGGAATAATACATACTAGATGAAATAAGTTACAGAT
...
562 AAAAGAGCCACAAACCCCTCAGTCGGCGGCCGACGCTTCCGATAGACTGCGTCCCGGG
612 617
622 TACCCGTAATCCCAATAAAGCCTCTGCTGTTTCATCCAAATGCT
...
901 GACCCGTGGTGAGACTGACGAGTCTGTGAACACCGCGCCGCAACCTGGAGAGCTCCAG
951 956
961 GGACTTTGGGGGCCGTTTGTGTGGCCGACCTGAGGAAGGAGTTCG
...
1081 TGAATTTTGGCTTGGGTTTGGAAACGAGCGCGCTTGTCTGCTGCGAGCGCTGCAG
1131 1136
|-----|
```

```
1141  CATCGTTCTGTGTGTCCTCTGTCTGACGTGTCTTCTGTAATTGTCTGAATAATTAGG
1142 1147
...
1206  TTACCACTCCCTTAAGTTTGAACCTTAGTCACTGGAAAGATGTGAGGCGGATCGCTCA
1256 1261
1266  ACCAGTCGGTAGATGTCAAGAAGAGACGTGGTTACCTTCTGCTGTGAGAAATGGCCAA
1290 1295
1326  CCTTTAAGTCGGAGGCGCGGAGACGACCTTTAACCGAGACCTCAACCCAGTTA
1331 1336
1386  A
...
1557  GTTGACCCCGCCTCGATCTCTCCTTAATCCAGCCCTCACTCTCTTAGCGCCGGA
1607 1612
1617  TTCCGATCTGATCAAGACAGACAGATGAGGATCGTTTCCCATGATTTGAACAAGATGATTG
1646 1651
1677  CACGAGGTCTCCGGCCGCTTGGG
...
1733  ACACACAATCGGCTGTCTGTATGCCCGCGGTGTCCGGCTGTACAGCGAGGCGCGCCGCT
1783 1788
1793  TCTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAATGTGACGACGAGCGAGCCCG
...
1853  GCTATCGTGGCTGCCACGACGAGGCGTTCCCTTGGCAGCTGTGCTGACGCTTGTCACTGA
1875 1880 1899 1904
1913  AGCGGAAGGAGCTGGCTGTATTGGGGGAAGTCCGGGGCA
...
2170  CCAGGCTCAAGGCGCGCATGCCGACGCGAGATCTGTGTGACCAATGCGCATGCGGCT
2220 2225
2230  GCTTGCCGAATATCATGTGTGAAATGCGCCGCTTTTCTGATTGATGACTGTGCGCGC
...
2290  TGGGTGTGGCGGACCGCTATCAGACATAGCGTGGCTACCCGATATTGCTGAAGAC
2318 2323
2350  TTGGCGGCGAATGGCTGACCGCTTCTCTGTGTTAGGATGCGGCTCCCGGATTGCG
2391 2396
2410  AGCGCATCGGCTTCTATCGGCTTCTTACGAGTTCTTCTGAGCGGAGCTCTGGGTTTCA
2415 2420 2425 2430
2470  AATGACCGACCAAGCGACCAACCTGCAATCAGAGATTGTGATTCCACCGCGGCTT
2482 2487 2485 2490
2530  CTATGAAGGTTGGCTTTCGGAATCGTTTTCGGGACGCGGCTGATGATCTCCAGCG
2552 2557 2564 2569
2590  CGGGATCTCATGCTGAGCTTCTTCCGCCACCCCGGGCTGATGCCCTCGGAGTTGTT
2592 2597
2650  CAGCTGTCCCTGAGGCTGAGGCTGAGGCTGCGGAGTTCTACCGGACGTCAAAATCCGTCG
2659 2674
2710  CATCCAGGAACACAG
...
2753  GAATGCAAGAGTGGGAGGACGATGGCCGCTTTGTGTGACCCGAGCGGAGGCTCTG
2803 2808
2813  CGCCTGATACAGAAAGAAATGCTTCAGGCACTCATGATGTGTCTTCCGTTTCCGC
...
2873  CTGAGTCACTGCGTGTGATGAGCGCTGGCGCTGTCTGCGGACGGAGCTGTACACA
2894 2899 2900 2905
2933  CCCACTGAGGGCGCTGACGCGCTGCAGAGGCGGAGCTGCAGAACTGCTCCAAAGGACCTC
2950 2955
2993  AAGCTTCCGAG
...
3183  AAGGGGGAAATGAAGACCCACCTGTAGTTTGGCAAGTACGTTAAGTAAAGCAATTT
3233 3238
3243  TGCAAGGCAATGAATAATACATAACTGAGAAATAGAAATTGAGAT
...
3589  GCGCTTGTCTCCCGAGCTCAATAAAGAGCCACAAACCCCTCACTGCGGCGCACTC
3639 3644
3649  CTCCGATTGACTGAGTGCAGCGCGGTACCCGATCAATTAACCTCTTGCAGTTGAT
3663 3668
3709  CCGACTTGTG
...
3967  CGTCAGGGGAGTGTGGCGGAGTGTGGGCGCAGCATGACCCAGTCAAGTACGATGCGG
4017 4022
4027  GAGTGTATCTGCGTTAATATGCGGATCAGAGCAGATTGTACTG
...
4084  TATGCGGTGTAATAATACCGACAGAGTGGTAAGGAGAAATACCGCATCAGGCGCTTTC
4134 4139
4144  CGCTTCCTGCTCACTGACTCGCTGCGGCTGCGTCTGCGGCTGCGGCGGAGGATATCAGC
4173 4180
4204  TCACTCAAGGCGGTAATACGTTATC
...
|-----|
```

4265 TGAGCAAAAGCGCAGCAAAAGCGGAGACCGTAAAAAGCGCGCTGTGGCTTTTC 4315 4320
4325 CATAGGCTCCGCCCCCTGAGAGACATCACAAAAATCGACGCTCAAGTACAGAGTGGCA 4362 4367
4385 AACCCGACAGGACTATAAGATACAGGCGTTCCCTGGAGACTCCCTGTCGCTCT 4411 4416
4445 CCTGTTCCGACCTGCGCGTTACCGGATACCTGTCGCTTCTCCTTCGGAGAGCTG 4504
4505 GCGCTTTCTCATAGCTCAGCGCTGAGTATCTAGTTCGTTGCTGCTCCAG 4550 4555
4565 CTGGGCTGTGTGACGAAACCCCGCTTACGCCGCTGCGCTTATCCGGTACTAT 4623
4625 CGTCTGAGTCCAAACCGGTAAAGACAGACTTATCGCCACTGGCAGCAGCCACTGTAA 4628 4657 4662
4685 AGGATTAGCAGACGAGGTATGTAGCGG
4890 GCAGAAAAAAGCATCTCAAGAAAGATCCTTTGATCTTTCTACGGGGCTGACGCTCAGT 4940 4945
4950 GGAACGAAAACTCACGTTAAGGATTTTGGTCAATGAGATTATCAAA
5123 CTATTTGCTCATCCATAGTTGCTGACTCCCGCTCGTAGATAA
5328 ATTGTTCCGCGGAAGTAGTAGTGTTCGCCAGTTAATAGTTTGGCAGACGTTGTG 5378 5383
5388 CCAATGCTGCAAGCATCGTGTGTCAACGCTGTGTTGGTATGCGTTCAATTCAGCTCCG 5419 5424
5448 GTTCCCAAGCATCAAGCGAGTTACATGATCCCATGTTGTGCAAAAAAGCGTTAGCT 5454 5459
5508 CCTTCGCTCCTCCGATCGTTGTCAGAGTAAGTTGGCCGAGTGTATCACTATGTTA 5522 5527
5568 TGGCAGCACT
5617 TTCTGTACTGCTGAGTACTCAACCAAGTCTTCTGAGAAATAGTGTATGCGGGCAGCCGAG 5667 5672
5677 TTGCTTTGGCCGCGTCAACAGCGGATTAATACCGGCCACATATGACAGAACTTTAAAGT

5689 5694
5737 GCTTCATCTTGGAAAGCTTCTTCGGGGCGCAAAACTCTCAAGATCTTACCCTTTGAG 5751 5756
5797 ATCCAGTTTCG
5808 TGTAAACCACTGTGACCACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTCTG 5858 5863
5868 GTTGACCAAAACGAGAGCGCAAAATGCCCAAAAAGGAAATAGGCGACACGCAAT 5914 5919
5928 GTTGATATCTCATCTCTCTTTTTCATATATTATGAAGCA
6021 AATATAACAATAGGGGTTCCGCGCATTTCCCGCAAAAGTCCACCTGACGCTTAAG 6071 6076
6081 AAACCATTAATTATCATGACATTAACCTATAAATAAGCGATATCAC
1 match found in sequence:
aav83189 : Granulocyte macrophage colony stimulating factor (GM-CSF) gene.
(from "mycobacterieng.seq")
TOIG of: aav83189 check: 6618 from: 1 to: 435
ID AAV83189 standard; DNA: 435 BP.
XX AC AAV83189;
XX DT 02-MAR-1999 (first entry)
XX DE Granulocyte macrophage colony stimulating factor (GM-CSF) gene.
XX KW Mouse: immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW Immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KW Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
XX KW human immunodeficiency virus; ds.
OS Homo sapiens.
XX PN WO9844788-A2.
XX PD 15-OCT-1998.
XX PF 09-APR-1998; 98MO-US06944.
XX PR 09-DEC-1997; 97US-0069163.
XX PR 09-APR-1997; 97US-0838702.
XX PR 01-MAY-1997; 97US-0848760.
XX PA (CHAN/) CHANG L.
XX PI Chang L;
XX DR WPI; 1999-024005/02.
XX PT Use of immunodeficient mice comprising human cells - particularly
PT SCD/Deige mice comprising human immune cells for evaluating
PT vaccines against cancers or human pathogens, e.g. HIV
XX PS Example 1b; Page 103-104; 1544p; English.

CC Immunodeficient mice comprising human cells can be used for
CC exposure to human pathogens and/or their components or human
CC tumour cells and human peripheral blood lymphocytes. Also claimed
CC is a vaccine comprising a cell modified to express an antigen and
CC an immune-modulating protein, this is preferably an expression
CC vector comprising a polynucleotide sequence that encodes the
CC antigen and immune-modulating protein. Such vectors can be used
CC to treat a subject having a tumour by transferring the expression
CC vector into the tumour so that the antigen and the immune-modulator
CC are expressed by at least the tumour. The methods can be used for
CC producing and evaluating vaccines including cancer vaccines and
CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC Mycobacterium, Listeria or Plasmodium. This sequence is an
CC intermediate, retroviral vector derived from pLNL6, a vector approved
CC for clinical use in the United States. pLNL is essentially pLNL6
CC digested with ClaI and BclI to remove cloning sites and the pLNL6
CC internal Sneo gene. These were replaced with a polylinker. A neo
CC gene was then inserted under the transcriptional control of the SV40
CC enhancer/promoter to create the vector pLSN. Two primers (AAV83185,
CC AAV83186) were used to amplify the granulocyte macrophage colony
CC stimulating factor (GM-CSF) gene (AAV83189) for its insertion into
CC pLSN.
XX
SQ Sequence 435 BP; 99 A; 140 C; 113 G; 83 T; 0 other;
AAV83189 Length: 435 March 5, 2002 14:17 Type: N Check: 6618 ..
Found using 'seq2-3' (pappu403.key)

69 CAGGCCGACGACGCCCTTGAGAGCATGTGATGCCATCCAGAGGCCGCTCTCT
119 124
129 GAACTGAGTAGAGACACTGCTGCTGAGATGATGAATGAACAGTAGAA
13 matches found in sequence:
aav83192; Efilalpa enhancer/promoter element.
(from "mycobacterieng.seq")
TOIG Of: aav83192 check: 8534 from: 1 to: 1451

ID AAV83192 standard; DNA; 1451 BP.
XX
AC AAV83192;
XX
DT 02-MAR-1999 (first entry)
XX
DE Efilalpa enhancer/promoter element.
XX
KW Mouse; Immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW Immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KW Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
KW human immunodeficiency virus; ss.
XX
OS Homo sapiens.
XX
PN WO844788-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-US06944.
XX
PR 09-DEC-1997; 97US-0069163.
PR 09-APR-1997; 97US-0838702.
PR 01-MAY-1997; 97US-0848760.
XX
PA (CHAN/) CHANG L.
XX
PI Chang L;
XX

DR WPI: 1999-024005/02.
XX
XX Use of immunodeficient mice comprising human cells - particularly
PR SCRP/beige mice comprising human immune cells for evaluating
PR vaccines against cancers or human pathogens, e.g. HIV
XX
XX Example 10a; Page 106-107; 154pp; English.
PS
XX
XX Immunodeficient mice comprising human cells can be used for
CC exposure to human pathogens and/or their components or human
CC tumour cells and human peripheral blood lymphocytes. Also claimed
CC is a vaccine comprising a cell modified to express an antigen and
CC an immune-modulating protein, this is preferably an expression
CC vector comprising a polynucleotide sequence that encodes the
CC antigen and immune-modulating protein. Such vectors can be used
CC to treat a subject having a tumour by transferring the expression
CC vector into the tumour so that the antigen and the immune-modulator
CC are expressed by at least the tumour. The methods can be used for
CC producing and evaluating vaccines including cancer vaccines and
CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC Mycobacterium, Listeria or Plasmodium. Two primers (AAV83193,
CC AAV83194) were used to amplify the Efilalpa promoter/enhancer element.
CC The Efilalpa promoter/enhancer is abundantly transcribed in a broad
CC range of cell types and was used in the construction of vectors in
CC which high levels of expression of therapeutic genes were desired.
XX
SQ Sequence 1451 BP; 252 A; 409 C; 458 G; 332 T; 0 other;
AAV83192 Length: 1451 March 5, 2002 14:17 Type: N Check: 8534 ..
Found using 'seq2-3' (pappu403.key)

106 AGTAATTCATCAAAAGGACTCGCCCTCTTGAGGATCCAGGACGCTGCTAAC
156 161
166 TCCCACTAACGTAGAACCCAGAGATCGCTGCCGCCGCCCTCACC GCCGCTCG
189 194
226 TCATCAGTAGAGTGAGAGAGGATGCGTCCGCTCCGCTCCGCTAGTGGCAGAC
286 GCACTGCGCCGACAGTCCCGGAGAAAGTTGGGGAGGGGTCGCAATTGAACCGGTCT
290 295
392 GCTCGGCTTTTTCGAGGGTGGGGAGAACCCGTAATATAGTAGAGTAGTCCCGTGA
442 447 451
452 ACGTCTTTTTCGACAGGGTTTCCCGCAGACACAGTAGTCCGTGTGG
456
...
637 TTGCGCTTAAGAGAGCCCTTCGCTCGCTTGTAGTTGAGGCTTGCCCTGGCGCTGGG
687 692
697 CCCCCGGTGGAATCTGTGGACACTTCGGGCTGTCTGCTCTTTCGATAGTCTCT
757 AGCCATTAAATTTTGTATGACCTGCTGCGACCTTTTTCGGAAGATAGTCTTGT
787 792
817 AAATGGGGCCAGATCTGCACACTGTAATTTTGGGTCGCGGGGCGGCGGCGG
869 874

```
877      GGGCCGTGGCTCCAGCGACATGTTGGGCGAGGCGGCGCTGCGAGC
...
951      AGTCTCAAGCTGGCGCGGCTGCTCTGTGGTCCCTGCGCGCGCGCTGTATGCGCCGC
                                     |-----|
                                     1001 1006
1011      CTTGGCGGCGCAAGCTGGCGCGCGCTGGCACCAGTTGGGTGAGCGGA
1057      AAGATGGCGCGCTTCCGCGCGCTGCTGCGAGGAGCTCAAAATGAGAGACGCGCGCTCGGG
                                     |-----|
                                     1107 1112
1117      AGAGCGGCGGCGGTGAGTCAACCCACACAAAGAAAGGCGCTTCCGCTCAGCCGCGCGC
                                     |-----|
                                     1172
1177      |-----|
          |-----|
          TTTGAGTGAAGTCCAGCGAGTACCGGCGCGCTGCGAGCAGCTGATTTCTGAGCTT
          1177      1201 1206
1237      TTTGAGTGAAGTCCAGCGAGTACCGGCGCGCTTATGCGATGAGTTTCCCGACAC
          1246 1251
1297      TGAGT
...
-----
2 matches found in sequence:
aav83199 : Interleukin (IL)-12B cDNA.
(from "mycobacterieng.seq")
TOIG of: aav83199 check: 2387 from: 1 to: 987

ID      AAV83199 standard; DNA: 987 BP.
XX
AC      AAV83199:
XX
DT      02-MAR-1999 (first entry)
XX
DE      Interleukin (IL)-12B cDNA.
XX
KW      Mouse: immunodeficient; pathogen: tumour; lymphocyte: antigen;
KW      immunomodulator; vector; vaccine: cancer; HIV; Leishmania;
KW      Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
KW      human immunodeficiency virus; ds.
XX
OS      Homo sapiens.
XX
PN      WO9844788-A2.
XX
PD      15-OCT-1998.
XX
PF      09-APR-1998; 98WO-US06944.
XX
PR      09-DEC-1997; 97US-0069163.
PR      09-APR-1997; 97US-0838702.
PR      01-MAY-1997; 97US-0848760.
XX
PA      (CHAN/) CHANG L.
XX
PI      Chang L;
XX
DR      WPI: 1999-024005/02.
XX
PT      Use of immunodeficient mice comprising human cells - particularly
XX      SCTD/beige mice comprising human immune cells for evaluating
XX      vaccines against cancers or human pathogens, e.g. HIV
XX
PS      Example 10(c)(1); Page 110; 154pp; English.
XX
```

```
CC      Immunodeficient mice comprising human cells can be used for
CC      exposure to human pathogens and/or their components or human
CC      tumour cells and human peripheral blood lymphocytes. Also claimed
CC      is a vaccine comprising a cell modified to express an antigen and
CC      an immune-modulating protein, this is preferably an expression
CC      vector comprising a polynucleotide sequence that encodes the
CC      antigen and immune-modulating protein. Such vectors can be used
CC      to treat a subject having a tumour by transferring the expression
CC      vector into the tumour so that the antigen and the immune-modulator
CC      are expressed by at least the tumour. The methods can be used for
CC      producing and evaluating vaccines including cancer vaccines and
CC      vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC      Mycobacterium, Listeria or Plasmodium. Two primers (AAV83197,
CC      AAV83198) were used to amplify the IL-12B for insertion into
CC      expression vectors of the invention.
XX
SQ      Sequence 987 BP; 264 A; 245 C; 257 G; 221 T; 0 other:
AAV83199 Length: 987 March 5, 2002 14:17 Type: N Check: 2387
Found using 'seq2-3' (pappu403.key)
...
363      AGAACCCAAAATAGACTTTCTAAGATGCGAGGCCAAGAAATATTCTGGAGCTTTCAC
          |-----|
          413 418
423      CTGCTGTGGCTGACGACATCAATCAATTTGACATTGAGTGTCAAAAGCAGCAGAGG
          |-----|
          435 440
483      CTCTTCTG
...
-----
4 matches found in sequence:
aav83200 : Sequence ID number 25 from WO9844788.
(from "mycobacterieng.seq")
TOIG of: aav83200 check: 6508 from: 1 to: 2097

ID      AAV83200 standard; DNA: 2097 BP.
XX
AC      AAV83200:
XX
DT      05-MAR-1999 (first entry)
XX
DE      Sequence ID number 25 from WO9844788.
XX
KW      Mouse: immunodeficient; pathogen: tumour; lymphocyte: antigen;
KW      immunomodulator; vector; vaccine: cancer; HIV; Leishmania;
KW      Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
KW      human immunodeficiency virus; ss.
XX
OS      Synthetic.
XX
PN      WO9844788-A2.
XX
PD      15-OCT-1998.
XX
PF      09-APR-1998; 98WO-US06944.
XX
PR      09-DEC-1997; 97US-0069163.
PR      09-APR-1997; 97US-0838702.
PR      01-MAY-1997; 97US-0848760.
XX
PA      (CHAN/) CHANG L.
XX
PI      Chang L;
XX
DR      WPI: 1999-024005/02.
XX
PT      Use of immunodeficient mice comprising human cells - particularly
XX
```

PT SCRD/beige mice comprising human immune cells for evaluating
 XX vaccines against cancers or human pathogens, e.g. HIV
 XX
 XX Disclosure: Page 111-112; 154pp; English.
 XX
 CC The invention relates to immunodeficient mice comprising human cells
 CC that can be used for exposure to human pathogens and/or their components
 CC or human tumour cells and human peripheral blood lymphocytes. Also
 CC claimed is a vaccine comprising a cell modified to express an antigen and
 CC an immune-modulating protein, this is preferably an expression
 CC vector comprising a polynucleotide sequence that encodes the
 CC antigen and immune-modulating protein. Such vectors can be used
 CC to treat a subject having a tumour by transferring the expression
 CC vector into the tumour so that the antigen and the immune-modulator
 CC are expressed by at least the tumour. The methods can be used for
 CC producing and evaluating vaccines including cancer vaccines and
 CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
 CC Mycobacterium, Listeria or Plasmodium. Mutations in the tat and
 CC nef genes of HIV provide live attenuated HIV vaccines. These
 CC tat-nef clones are replication competent and may be employed as
 CC proviral DNA vaccines or as a vaccine comprising viral particles.
 XX
 SQ Sequence 2097 BP; 545 A; 491 C; 579 G; 482 T; 0 other;
 AAV83200 Length: 2097 March 5, 2002 14:17 Type: N Check: 6508
 Found using 'seq2-3' (pappu403.key)
 ...
 89 TGTCGAGCATGAGCCACTAAGTGCAGAGTTCGCGACCATATGAAAGCGTCATTC
 |-----|
 139 144
 149 CATCCGATGTCGCCAGTGTGCTTGTGTGAGAAAGCCTCTACTCT
 ...
 1139 ACAGTGTAGGGAATAATAGAGTGTATCAGCAGACGACCGAAGACTGCATGCCCAAGA
 |-----|
 1189 1194
 1199 TCATGAATGGAGAGAGCTGATGCCATGAGCTTGATGGAGGTTTGT
 ...
 1250 TAGCGGCGAAGTGTGTGTGTCCTGCTTGGCAGAAACTACATTAAGAGCGATAATT
 |-----|
 1300 1305
 1310 GTGAGATACACGACGAGGATATTTTGTGTAGCAGTGTGAA
 ...
 1562 GCCTAAACCTGTGTGAACCAACAAGAGGATACAGGCTACACAGCGCTTCA
 |-----|
 1612 1617
 1622 GGTGTCTGTGAGAGAGGAGATGTGCGCTTTGTGAAACACACAGAC
 ...

 1 match found in sequence:
 aav83201; Wild type sequence of HIV tat-B gene fragment.
 (from "mycobacterieng.seq")
 FIG off: aav83201 check: 1040 from: 1 to: 140
 ID AAV83201 standard; DNA; 140 BP.
 XX
 XX AAV83201;
 AC
 XX
 XX 02-MAR-1999 (first entry)
 DT
 XX

DE Wild type sequence of HIV tat-B gene fragment.
 XX
 XX Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
 KW immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
 KW Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
 KW human immunodeficiency virus; ss.
 XX
 OS Human immunodeficiency virus.
 XX
 PN W09844788-A2.
 XX
 PD 15-OCT-1998.
 XX
 XX 09-APR-1998; 98WO-US06944.
 PF
 XX
 PR 09-DEC-1997; 97US-0069163.
 PR 09-APR-1997; 97US-0838702.
 PR 01-MAY-1997; 97US-0848760.
 XX
 PA (CHANV) CHANG L.
 XX
 PI Chang L;
 XX
 DR WPI: 1999-024005/02.
 XX
 PT Use of immunodeficient mice comprising human cells - particularly
 PT SCRD/beige mice comprising human immune cells for evaluating
 PT vaccines against cancers or human pathogens, e.g. HIV
 XX
 XX Example 20(a); Figure 12; 154pp; English.
 XX
 CC Immunodeficient mice comprising human cells can be used for
 CC exposure to human pathogens and/or their components or human
 CC tumour cells and human peripheral blood lymphocytes. Also claimed
 CC is a vaccine comprising a cell modified to express an antigen and
 CC an immune-modulating protein, this is preferably an expression
 CC vector comprising a polynucleotide sequence that encodes the
 CC antigen and immune-modulating protein. Such vectors can be used
 CC to treat a subject having a tumour by transferring the expression
 CC vector into the tumour so that the antigen and the immune-modulator
 CC are expressed by at least the tumour. The methods can be used for
 CC producing and evaluating vaccines including cancer vaccines and
 CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
 CC Mycobacterium, Listeria or Plasmodium. Mutations in the tat and
 CC nef genes of HIV provide live attenuated HIV vaccines. These
 CC tat-nef clones are replication competent and may be employed as
 CC proviral DNA vaccines or as a vaccine comprising viral particles.
 XX
 SQ Sequence 140 BP; 43 A; 28 C; 35 G; 34 T; 0 other;
 AAV83201 Length: 140 March 5, 2002 14:17 Type: N Check: 1040
 Found using 'seq2-3' (pappu403.key)
 ...
 1 AGCGTTACTCGACAGAGAGACAGAAATGAGCCAGTAGATCTAGACTAGAGC
 |-----|
 2 7

 1 match found in sequence:
 aav83204; Primer for amplifying HIV gag gene.
 (from "mycobacterieng.seq")
 FIG off: aav83204 check: 8789 from: 1 to: 28
 ID AAV83204 standard; DNA; 28 BP.
 XX
 XX AAV83204;
 AC
 XX
 XX 02-MAR-1999 (first entry)
 DT
 XX
 DE Primer for amplifying HIV gag gene.
 XX

```
KW Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
KM Immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KM Mycobacterium; Listeria; Plasmidium; retrovirus; evaluation;
KM human immunodeficiency virus; PCR primer; ss.
XX
OS Synthetic.
OS Human immunodeficiency virus.
XX
PN WO9844788-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98MO-US06944.
XX
PR 09-DEC-1997; 97US-0069163.
PR 09-APR-1997; 97US-0838702.
PR 01-MAY-1997; 97US-0848760.
XX
PA (CHAN/) CHANG L.
XX
PI Chang L.
XX
DR WPI; 1999-024005/02.
XX
PT Use of immunodeficient mice comprising human cells - particularly
PT SCID/beige mice comprising human immune cells for evaluating
PT vaccines against cancers or human pathogens, e.g. HIV
XX
PS Example 20(b)(1); Page 87; 154pp; English.
XX
CC Immunodeficient mice comprising human cells can be used for
CC exposure to human pathogens and/or their components or human
CC tumour cells and human peripheral blood lymphocytes. Also claimed
CC is a vaccine comprising a cell modified to express an antigen and
CC an immune-modulating protein, this is preferably an expression
CC vector comprising a polynucleotide sequence that encodes the
CC antigen and immune-modulating protein. Such vectors can be used
CC to treat a subject having a tumour by transferring the expression
CC vector into the tumour so that the antigen and the immune-modulator
CC are expressed by at least the tumour. The methods can be used for
CC producing and evaluating vaccines including cancer vaccines and
CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC Mycobacterium, Listeria or Plasmidium. Two primers (AAV83204,
CC AAV83205) were used to amplify the HIV gag gene for insertion into
CC the expression vectors pHP-1 (AAV83106). pHP-1 comprises all of the
CC structural genes of HIV but none of the genes required for
CC packaging of the virus into viral particles. It thus mimics HIV
CC infection when the viral proteins are expressed.
XX
SQ Sequence 28 BP; 5 A; 8 C; 11 G; 4 T; 0 other;
AAV83204 Length: 28 March 5, 2002 14:17 Type: N Check: 8789 ..
Found using 'seq2-3' (pappu403.key)
1 CGGATCCACCATGATGTCGAGAGCTC
1 23 28
-----
53 matches found in sequence:
aav83206; Vector pHP-1 comprising all HIV structural genes.
(from "mycobacterieng.seq")
TOIG of: aav83206 check: 4258 from: 1 to: 12494
ID AAV83206 standard; DNA: 12494 BP.
XX
AC AAV83206;
XX
DT 02-MAR-1999 (first entry)
XX
DE Vector pHP-1 comprising all HIV structural genes.
XX
KW Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
```

```
KW Immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KM Mycobacterium; Listeria; Plasmidium; retrovirus; evaluation;
KM human immunodeficiency virus; ds.
XX
OS Synthetic.
OS Human immunodeficiency virus; ds.
XX
PN WO9844788-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98MO-US06944.
XX
PR 09-DEC-1997; 97US-0069163.
PR 09-APR-1997; 97US-0838702.
PR 01-MAY-1997; 97US-0848760.
XX
PA (CHAN/) CHANG L.
XX
PI Chang L.
XX
DR WPI; 1999-024005/02.
XX
PT Use of immunodeficient mice comprising human cells - particularly
PT SCID/beige mice comprising human immune cells for evaluating
PT vaccines against cancers or human pathogens, e.g. HIV
XX
PS Example 20(b)(1); Page 114-122; 154pp; English.
XX
CC Immunodeficient mice comprising human cells can be used for
CC exposure to human pathogens and/or their components or human
CC tumour cells and human peripheral blood lymphocytes. Also claimed
CC is a vaccine comprising a cell modified to express an antigen and
CC an immune-modulating protein, this is preferably an expression
CC vector comprising a polynucleotide sequence that encodes the
CC antigen and immune-modulating protein. Such vectors can be used
CC to treat a subject having a tumour by transferring the expression
CC vector into the tumour so that the antigen and the immune-modulator
CC are expressed by at least the tumour. The methods can be used for
CC producing and evaluating vaccines including cancer vaccines and
CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC Mycobacterium, Listeria or Plasmidium. pHP-1 requires all of the
CC structural genes of HIV but none of the genes required for
CC packaging of the virus into viral particles. It thus mimics HIV
CC infection when the viral proteins are expressed.
XX
SQ Sequence 12494 BP; 4121 A; 2467 C; 3024 G; 2882 T; 0 other;
AAV83206 Length: 12494 March 5, 2002 14:17 Type: N Check: 4258 ..
Found using 'seq2-3' (pappu403.key)
...
39 AGTTCGGCGTTACATACTTACGTAATGCGCCGCTGCGTACGCCGCCACAGACCCCC
1 89 94
99 GCCCATTTGACGCTCAATATGATGATGTTCCATAGTACGCAATATGAGGACCTTCCATT
1 106 111 136 141
159 GAGCGTAATGGAGATTGTTTGGACCAAAATCAACGGGACCTTCCAAATATGTCG
159 164
...
349 GCTCTGTGCTAAGGAGAACCCACTGCTTAAGCCTCAATTAAGCTTTGCGCCGGT
1 399 404
409 GGATCAAGACCGGTCGCTCAATTAAGGTGATTTCGTGCGATCCACGAGAGATGGGTGC
1 409
-----
```



```

9839 GTGTGTTCCAGTTTGGAAACAAAGATCCTATTAAAGAACGTGAGCTCCACGTCGAAG
      9889 9894
9899 GCGCAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCCTAATCAAGTT
      9919 9924
9959 TTTTGGGTCGAGGTG
...
10031 GAAAGCCGGCGAAGCTGCGAGAGAAAGAAAGAAAGAAAGCGAGCGGCGCTAGGG
      10081 10089
10091 ---|
      10094 CCGTGGCAAGCTGAGCGGTGACGTCGCGGCTAACCAACACCGCGCGCTTA
...
10268 TGCTTCAATATATATGAAAAAGAGAGATATGAGTATTCACATTTCCGTGCGCCCTTA
      10318
10328 TTCCCTTTTTTGGGCGCATTTTGGCTTCCTGTTTTTGTCTACCCAGAAAGCGCTGTGAAG
      10374
10388 TAAAGATGCTGAAGATCAGTGGTGCACGAGTGGGTTCAC
...
10431 CGAATGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCGAAGACGTTTCC
      10481
10491 AATGATGAGCAGCTTTAAAGTCTGCTATGTGCGCGCGTATTTATCCGTAATTCAGTTC
      10543
10551 GCAAGAGCAACTCGTCGCGCCATACACTATCTCAGAAATGACTTGGTGTGATCTCAC
      10565
10611 AGTCACAGAA
...
10660 AGTCTGCCATTAACATGATGATTAACACTGCGGCCAATTACTTCTGACAAAGATCGGA
      10710
10720 GGAACGAAGAGCTAACCGCTTTTTCACACATGSGGGATCATGTAACCTCGCTTGAT
      10778
10780 ---|
      10783 CGTTGGGAACCGAGCTGAATGAAGCCATACCAAAAGCAGCGGTGACACAGATGCCCT
      10813
10840 GTAGCAATGGCAACAGTTGCGCAAACTATTAACTGGCACTACTTACTTACTGCTTCC
      10854
10900 CCGCAACAAT
...
11069 TTAATCTACGACGAGGAGTCAGGCACTATGATGTAACGAAATAGACAGATGCTGAGA
      11119

```

```

11129 TAGTGCCATCAGTATTAAAGCATTTGTAAGTCTGACAGCAAGTTTA
...
11242 TTGTATATCTCATGACCAAAATCCCTTAACGTGAGTTTTGTTCCACTGAGGCTCAGAC
      11292
11302 CCGGTAGAAAAGATCAAGAGATCTTCTTGATATCCTTTTTTCTGC
...
11525 CGCCTACATATACCTCGCTCTGCTAATCCTGTTTACCAGTGGCTGCTGCCAGTGGGGATAGT
      11575
11585 CGTGTCTTACCGGGTTGACTCAAGACGATAGTTTACCGGATTAAGGCGCACCGGCTCGGGCT
      11609
11645 GAACGGGGGCTTGTGCACACAGCCCACTTGGAGCGAAGACCTTACCCGAACCTAGAT
      11682
11705 ACCTACAGCGTGAAGTATGAGAAAGCGCCACGCTTCCCGAAGGAGAAAGCGGACAGT
      11728
11765 ATCCGGTAAAGCGGCGAGGTCGGAACAGAGAGCGCACAGGAGCTTCCAGGGGGAACG
      11821
11825 ---|
      11826 CTTGTAATCTTTATATGTCCTGTGCGGTTTCCCACTCTGACTTGACCTGCAATTTTGT
      11870
11885 GATGCTGTCAGGGGGGCGAGGCTATGAAAAACGACAGCAACGCGGCTTTTACGGT
      11917
11945 TCCTGGCCTTTTGTGCGCTTTGCTCA
...
12007 GATTAACCGTATTACCGCTTTGAGTGAAGCTGATACCGCTCGCCGACCCAGACGAG
      12057
12067 CGCAGCGAGTCAGTGAAGCGAAGAAAGCGGAGGCCCAATATAGCAAAACGCGCTCTCCCC
      12098
12127 GCGCGTTGGCCGATTCATTATGCAAGC
...
1 match found in sequence:
aav83207 ; Interleukin (IL)-12A cDNA.
(from "mycobacterieng.seq")
TOIG of: aav83207 check: 1669 from: 1 to: 762
ID AAV83207 standard; cDNA; 762 bp.
XX AC
XX AAV83207;
XX AC
XX 02-MAR-1999 (first entry)
XX DT
XX DE Interleukin (IL)-12A cDNA.
XX KW Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW immunomodulator; vector; vaccine; cancer; HIV; leishmania;
KW Mycobacterium; Listeria; Plasmidium; retrovirus; evaluation;

```

```

KW human immunodeficiency virus; ds.
XX
OS Homo sapiens.
XX
PN MO9844788-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-US06944.
XX
PR 09-DEC-1997; 97US-0069163.
PR 09-APR-1997; 97US-0838702.
PR 01-MAY-1997; 97US-0848760.
XX
PA (CHAN/) CHANG L.
XX
PI Chang L;
XX
DR WPI: 1999-024005/02.
XX
PT Use of immunodeficient mice comprising human cells - particularly
PT SCID/beige mice comprising human immune cells for evaluating
PT vaccines against cancers or human pathogens, e.g. HIV
XX
PS Example 10(c)(1); Page 108-109; 154pp; English.
XX
CC Immunodeficient mice comprising human cells can be used for
CC exposure to human pathogens and/or their components or human
CC tumour cells and human peripheral blood lymphocytes. Also claimed
CC is a vaccine comprising a cell modified to express an antigen and
CC an immune-modulating protein, this is preferably an expression
CC vector comprising a polynucleotide sequence that encodes the
CC antigen and immune-modulating protein. Such vectors can be used
CC to treat a subject having a tumour by transferring the expression
CC vector into the tumour so that the antigen and the immune-modulator
CC are expressed by at least the tumour. The methods can be used for
CC producing and evaluating vaccines including cancer vaccines and
CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC Mycobacterium, Listeria or Plasmodium. Two primers (AAV83197,
CC AAV83198) were used to amplify the IL-12B for insertion into
CC expression vectors of the invention.
XX
SQ Sequence 762 BP; 199 A; 211 C; 166 G; 186 T; 0 other:
AAV83207 Length: 762 March 5, 2002 14:17 Type: N Check: 1669 ..
Found using 'seq2-3' (pappu403.key)
...
80 TGTCCCTGCAGTGGCGGCTCAGCATGTGTCCAGCGCGGCGAGCTCTCTTGTGCTACCC |-----|
130 135
140 TGTCTCTCTGAGACACCTCAGTTTGCCAGAAACCTCCCGTGCGC
...
1 match found in sequence:
aav83208 ; B7-2 cDNA.
(from "mycobacterieng.seq")
TOIG off: aav83208 check: 992 from: 1 to: 972
ID AAV83208 standard; cDNA; 972 BP.
XX
XX AAV83208;
AC
XX
XX 02-MAR-1999 (first entry)
DT
XX
DE B7-2 cDNA.
XX
KW Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW immunomodulator; vector; vaccine; cancer; HIV; Leishmania;

```

```

KW Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
XX human immunodeficiency virus; ds.
XX
OS Homo sapiens.
XX
PN MO9844788-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-US06944.
XX
PR 09-DEC-1997; 97US-0069163.
PR 09-APR-1997; 97US-0838702.
PR 01-MAY-1997; 97US-0848760.
XX
PA (CHAN/) CHANG L.
XX
PI Chang L;
XX
DR WPI: 1999-024005/02.
XX
PT Use of immunodeficient mice comprising human cells - particularly
PT SCID/beige mice comprising human immune cells for evaluating
PT vaccines against cancers or human pathogens, e.g. HIV
XX
PS Example 1b; Page 104-105; 154pp; English.
XX
CC Immunodeficient mice comprising human cells can be used for
CC exposure to human pathogens and/or their components or human
CC tumour cells and human peripheral blood lymphocytes. Also claimed
CC is a vaccine comprising a cell modified to express an antigen and
CC an immune-modulating protein, this is preferably an expression
CC vector comprising a polynucleotide sequence that encodes the
CC antigen and immune-modulating protein. Such vectors can be used
CC to treat a subject having a tumour by transferring the expression
CC vector into the tumour so that the antigen and the immune-modulator
CC are expressed by at least the tumour. The methods can be used for
CC producing and evaluating vaccines including cancer vaccines and
CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC Mycobacterium, Listeria or Plasmodium. This sequence is an
CC intermediate retroviral vector derived from pNL6, a vector approved
CC for clinical use in the United States. pNL6 is essentially pNL6
CC digested with ClaI and BclI to remove cloning sites and the pNL6
CC Internal Syno gene. These were replaced with a polylinker. A neo
CC gene was then inserted under the transcriptional control of the SV40
CC enhancer/promoter to create the vector pLSN. Two primers (AAV83187,
CC AAV83188) were used to amplify the B7-2 cDNA (a ligand for
CC CD28/CTLA-4 proteins, co-stimulators for interleukin-2 driven
CC proliferation of T-cells) for its insertion into pLSN.
XX
SQ Sequence 972 BP; 304 A; 204 C; 194 G; 270 T; 0 other:
AAV83208 Length: 972 March 5, 2002 14:17 Type: N Check: 992 ..
Found using 'seq2-3' (pappu403.key)
...
524 AGTATGATGATTATTCAGAAATTCACAGATATATGACAGAACTGACAGTTTCCA |-----|
574 579
584 TCAGCTTGTCTGTTTTCATTCCCTGAGTGTACGAGCAATATGACCAT
...
15 matches found in sequence:
aav99751 ; M. luteus Rp-factor encoding gene.
(from "mycobacterieng.seq")
TOIG off: aav99751 check: 8139 from: 1 to: 758
ID AAV99751 standard; DNA; 758 BP.
XX

```

```

AC AAV99751;
XX
XX 26-MAR-1999 (first entry)
XX
XX M. luteus RP-factor encoding gene.
DE
XX RP-factor; resuscitate; latent cell; growth-stimulation; receptor;
KW convertase; vaccine; immunotherapy; prophylaxis; excipient; antibiotic;
KW antimicrobial; tuberculosis; food; pharmaceutical; culture additive;
KW bacteria; ss.
XX
XX Micrococcus luteus.
OS
XX
XX Key Location/Qualifiers
FH CDS 65..728
FT /*tag= a
FT /product= "RP-factor"
XX
XX MO9855624-A1.
XX
XX 10-DEC-1998.
XX
XX 03-JUN-1998; 98WO-GB01619.
XX
XX 27-MAY-1998; 98GB-0011221.
XX 04-JUN-1997; 97GB-0011389.
XX
XX (UYMA-) UNIV WALES.
XX
XX Kaprelyants AS, Kell DB, Mukamolova GV, Young DI;
XX Young M;
XX
XX WPI: 1999-070270/06.
XX P-PSDB; AAM95417.
XX
XX New bacterial resuscitation factors - useful for stimulating latent
XX bacteria and growth, particularly for use as antimicrobials, also
XX PT vaccines comprising bacteria with mutations in resuscitation factor
XX genes
XX
XX Examples: Fig 2A; 76pp; English.
XX
XX This represents the sequence of the M. luteus RP-factor encoding gene.
XX CC RP-factors are substances that can resuscitate dormant, moribund or
XX CC latent cells, possibly also having growth-stimulating activity. Host
XX CC cells containing a vector comprising the RP-factor or its receptor
XX CC encoding nucleic acid can be used for the recombinant production of the
XX CC RP-factor. RP-factors, their receptors or convertases, antibodies
XX CC (against the RP-factors and RP-factor receptors or convertases),
XX CC antagonists and agonists, are useful in vaccines and for immunotherapy,
XX CC diagnosis and prophylaxis. They are also useful as excipients, generally
XX CC as antimicrobials and especially for infections associated with latency.
XX CC They can be used as potentiators of antibiotics such as isoniazid.
XX CC streptomycin etc., in treatment of tuberculosis. RP-factor can also be
XX CC used: to determine the microbiological quality of foods, pharmaceuticals,
XX CC medical products; as culture additive for bacteria; to stimulate growth
XX CC and/or to resuscitate microorganisms and to produce libraries of
XX CC biomolecules and microorganisms (which may then be screened for useful
XX CC products). Probes complementary to the RP-factor nucleic acid are used
XX CC to identify and clone other RP-factor genes. The inventions may also be used
XX CC to prevent bacterial resuscitation. Breaking dormancy with RP-factor
XX CC facilitates detection, culture and enumeration of many bacteria.
XX
XX Sequence 758 BP; 116 A; 297 C; 252 G; 93 T; 0 other;
S0
AAV99751 Length: 758 March 5, 2002 14:19 Type: N Check: 8139
Found using 'seq2-3' (pappu403.key)
1
ACCAAGAGAGACGACCGCGTGCTCGCGCCCGCATGACGAGACTCGCCATGG
13 18
|-----|
|-----|

```

```

61 ACACATGACTCTTTCACCACTTCGCGCACCCGCTCCGCGCTGCGACCGCTCGATCG
117
121 -|-----|
TGGCGGGCATGACCTTCGCGCGCGCGCGCGCTTGAGGCTTCTCGCGCGCGCGCGCGCG
122 141 146
181 CCACGCTGACACCTG
...
382 GCGCGTGGCGCGCTGTGCTGCGCAGAGACTGGGCGCTGACCCAGCTGACGCGGACCGGTCG
432 437
442 -|-----|-----|-----|-----|
ACGTGACCGCCACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
447 452 468 473 496 501
502 AGTCCCGCGCGGACGAGAGCTGCGCGCGCGAGAGCGCGCTGCGCGGAGACAGCGCGCTGCTG
555 560
562 -|-----|-----|-----|-----|
CCGAGCGCGGACGACCTGCTGCTCACTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
563 576 581 605 610
563 579 584
622 ACAGGTGAGAGGTGGCTGAGACCGCGCTTACGAGGCGCACAGAGCGCGCGCTGCTGAGCG
666 671 678
682 -|-----|-----|-----|-----|
CCGCGGTATCTACGTCGCGCAGAGAGCTGCTGCGCGCAGGCTGAGAGCGCTGACCGG
683 729 734
742 CCCCCCGACCGGTACC
-----
8 matches found in sequence:
aav99752: S. coelicolor partial RP-factor encoding DNA.
(from "mycobacterieng.seq")
TOIG of: aav99752 Check: 5276 from: 1 to: 299
ID AAV99752 standard; DNA; 299 BP.
XX
XX AC AAV99752;
XX
XX 26-MAR-1999 (first entry)
XX
XX S. coelicolor partial RP-factor encoding DNA.
XX
XX RP-factor; resuscitate; latent cell; growth-stimulation; receptor;
KW convertase; vaccine; immunotherapy; prophylaxis; excipient; antibiotic;
KW antimicrobial; tuberculosis; food; pharmaceutical; culture additive;
KW bacteria; ss.
XX
XX Streptomyces coelicolor.
XX
XX Key Location/Qualifiers
FH CDS 3..299
FT /*tag= a
FT /product= "partial RP-factor"
XX
XX MO9855624-A1.
XX
XX 10-DEC-1998.
XX
XX 03-JUN-1998; 98WO-GB01619.
XX

```

```

PR 27-MAY-1998; 98GB-0011221.
PR 04-JUN-1997; 97GB-0011389.
XX (UTWA-) UNIV WALES.
XX Kaprelyants AS, Kell DB, Mukamolova GV, Young DI;
XX Young M;
XX WPI: 1999-070270/06.
XX P-PSDB: AAW95418.
XX
XX New bacterial resuscitation factors - useful for stimulating latent
XX bacteria and growth, particularly for use as antimicrobials, also
XX vaccines comprising bacteria with mutations in resuscitation factor
XX genes
XX
XX Examples; Fig 2B; 76pp; English.
XX
XX The invention relates to RP-factors which are substances that can
XX resuscitate dormant, moribund or latent cells, possibly also having
XX growth-stimulating activity. Host cells containing a vector comprising
XX the RP-factor or its receptor encoding nucleic acid can be used for the
XX recombinant production of the RP-factor. RP-factors, their receptors or
XX convertases, antibodies (against the RP-factors and RP-factor receptors
XX or convertases), antagonists and agonists, are useful in vaccines and for
XX immunotherapy, diagnosis and prophylaxis. They are also useful as
XX excipients, generally as antimicrobials and especially for infections
XX associated with latency. They can be used as potentiators of antibiotics
XX such as isoniazid, streptomycin etc., in treatment of tuberculosis.
XX RP-factor can also be used: to determine the microbiological quality of
XX foods, pharmaceuticals, medical products; as culture additive for
XX bacteria; to stimulate growth and/or to resuscitate microorganisms and to
XX produce libraries of biomolecules and microorganisms (which may then be
XX screened for useful products). Probes complementary to the RP-factor
XX nucleic acid are used to identify and clone other RP-factor genes. The
XX inventions may also be used to prevent bacterial resuscitation. Breaking of
XX dormancy with RP-factor facilitates detection, culture and enumeration of
XX many bacteria. The present sequence represents a DNA encoding a partial
XX RP-factor from S. coelicolor.
XX
XX Sequence 299 BP; 41 A; 120 C; 106 G; 32 T; 0 other;
XX
AAV99752 Length: 299 March 5, 2002 14:19 Type: N Check: 5276
Found using 'seq2-3' (pappu403.key)
1 GGAATCCGACCGCGCGGTAACTGTGTGCGCGGACCGCACTGGGGCGCGCGAG
27 32 47 52
1 CGATGCGCGCGCGCGCGGTAACTGTGTGCGCGGACCGCACTGGGGCGCGCGAGT
77 82 99 104 105 110
61 CGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
121 CCAAGCGCACTGGGAGCGCAACCGCAAGCGCTACTAGGCGCGCGCGCGCGCGCGCG
181 GGTTCAGCTGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
192 197
241 GCGGCGGACAGATGCGCGGTGGCGGAGCGCTCTGTGCAAGGGATGTCCGCGTGG
252 257 265 270
-----
1 match found in sequence:
aav99757; M. tuberculosis RP-factor coding sequence amplifying primer.
(from "mycobacterieng.seq")
TOIG of: aav99757 Check: 4799 from: 1 to: 35
ID AAV99757 standard; DNA: 35 BP.
XX

```

```

AC AAV99757;
XX
XX 26-MAR-1999 (first entry)
XX
XX M. tuberculosis RP-factor coding sequence amplifying primer.
XX
XX RP-factor; resuscitate; latent cell; growth-stimulation; receptor;
XX convertase; vaccine; immunotherapy; prophylaxis; excipient; antibiotic;
XX antimicrobial; tuberculosis; food; pharmaceutical; culture additive;
XX PCR primer; ss.
XX
XX Synthetic.
XX Mycobacterium tuberculosis.
XX
XX W09855624-A1.
XX
XX 10-DEC-1998.
XX
XX 03-JUN-1998; 98WO-GB01619.
XX
XX 27-MAY-1998; 98GB-0011221.
XX 04-JUN-1997; 97GB-0011389.
XX
XX (UTWA-) UNIV WALES.
XX
XX Kaprelyants AS, Kell DB, Mukamolova GV, Young DI;
XX Young M;
XX WPI: 1999-070270/06.
XX
XX New bacterial resuscitation factors - useful for stimulating latent
XX bacteria and growth, particularly for use as antimicrobials, also
XX vaccines comprising bacteria with mutations in resuscitation factor
XX genes
XX
XX Examples; page 38; 76pp; English.
XX
XX The invention relates to RP-factors which are substances that can
XX resuscitate dormant, moribund or latent cells, possibly also having
XX growth-stimulating activity. Host cells containing a vector comprising
XX the RP-factor or its receptor encoding nucleic acid can be used for the
XX recombinant production of the RP-factor. RP-factors, their receptors or
XX convertases, antibodies (against the RP-factors and RP-factor receptors
XX or convertases), antagonists and agonists, are useful in vaccines and
XX for immunotherapy, diagnosis and prophylaxis. They are also useful as
XX excipients, generally as antimicrobials and especially for infections
XX associated with latency. They can be used as potentiators of antibiotics
XX such as isoniazid, streptomycin etc., in treatment of tuberculosis.
XX RP-factor can also be used: to determine the microbiological quality of
XX foods, pharmaceuticals, medical products; as culture additive for
XX bacteria; to stimulate growth and/or to resuscitate microorganisms and
XX to produce libraries of biomolecules and microorganisms (which may then
XX be screened for useful products). Probes complementary to the RP-factor
XX nucleic acid are used to identify and clone other RP-factor genes. The
XX inventions may also be used to prevent bacterial resuscitation. Breaking
XX of dormancy with RP-factor facilitates detection, culture and enumeration
XX of many bacteria. Primers AAV99757-58 are used for the PCR amplification
XX of the M. tuberculosis RP-factor coding sequence.
XX
XX Sequence 35 BP; 11 A; 7 C; 9 G; 8 T; 0 other;
XX
AAV99757 Length: 35 March 5, 2002 14:19 Type: N Check: 4799
Found using 'seq2-3' (pappu403.key)
1 ATCAGATTTCATATGAGCGACATCGATTGGGACCG
16 21
-----
1 match found in sequence:
aav99758; M. tuberculosis RP-factor coding sequence amplifying primer.
(from "mycobacterieng.seq")
TOIG of: aav99758 Check: 1142 from: 1 to: 29

```

ID AAV99758 standard; DNA: 29 BP.
XX
AC AAV99758;
XX
DT 26-MAR-1999 (first entry)
XX
DE M. tuberculosis RP-factor coding sequence amplifying primer.
XX
KW RP-factor: resuscitate; latent cell; growth-stimulation; receptor;
KW convertase; vaccine; immunotherapy; prophylaxis; excipient; antibiotic;
KW antimicrobial; tuberculosis; food; pharmaceutical; culture additive;
KW PCR primer; ss.
XX
OS Synthetic.
OS Mycobacterium tuberculosis.
XX
PN M09855624-A1.
XX
PD 10-DEC-1998.
XX
PE 03-JUN-1998; 98WO-GB01619.
XX
PR 27-MAY-1998; 98GB-0011221.
PR 04-JUN-1997; 97GB-0011389.
XX
XX (UYWA-) UNIV WALES.
XX
PI Kaprelyants AS, Kell DB, Mukamolova GV, Young DI;
PI Young M;
XX
DR WPI: 1999-070270/06.
XX
PT New bacterial resuscitation factors - useful for stimulating latent
PT bacteria and growth, particularly for use as antimicrobials; also
PT vaccines comprising bacteria with mutations in resuscitation factor
PT genes
XX
PS Examples; Page 38; 76pp; English.
XX
CC The invention relates to RP-factors which are substances that can
CC resuscitate dormant, moribund or latent cells, possibly also having
CC growth-stimulating activity. Host cells containing a vector comprising
CC the RP-factor or its receptor encoding nucleic acid can be used for the
CC recombinant production of the RP-factor. RP-factors, their receptors or
CC convertases, antibodies (against the RP-factors and RP-factor receptors
CC or convertases), antagonists and agonists, are useful in vaccines and
CC for immunotherapy, diagnosis and prophylaxis. They are also useful as
CC excipients, generally as antimicrobials and especially for infections
CC associated with latency. They can be used as potentiators of antibiotics
CC such as isoniazid, streptomycin etc., in treatment of tuberculosis.
CC RP-factor can also be used to determine the microbiological quality of
CC foods, pharmaceuticals, medical products; as culture additive for
CC bacteria; to stimulate growth and/or to resuscitate microorganisms and
CC to produce libraries of biomolecules and microorganisms (which may then
CC be screened for useful products). Probes complementary to the RP-factor
CC nucleic acid are used to identify and clone other RP-factor genes. The
CC inventions may also be used to prevent bacterial resuscitation. Breaking
CC dormancy with RP-factor facilitates detection, culture and enumeration
CC of many bacteria. Primers AAV99757-58 are used for the PCR amplification
CC of the M. tuberculosis RP-factor coding sequence.
XX
SQ Sequence 29 BP; 4 A; 14 C; 5 G; 6 T; 0 other;
AAV99758 Length: 29 March 5, 2002 14:19 Type: N Check: 1142 ..
Found using 'seq2-3' (pappu403.key)

1 CGCAGATCCCGTCATCGTCGTCGTCGC
16 21

13 matches found in sequence:

aa211303 : M. vaccae antigen 85A encoding DNA.
(from "mycobacter_n-seg")
TOIG of: aa211303 check: 182 from: 1 to: 1211
ID AA211303 standard; DNA: 1211 BP.
XX
AC AA211303;
XX
DT 25-OCT-1999 (first entry)
XX
DE M. vaccae antigen 85A encoding DNA.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN M09932634-A2.
XX
PD 01-JUL-1999.
XX
PE 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095885.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI P-PSDB; AAV14858.
XX
DR WPI: 1999-430163/36.
XX
PS P-PSDB; AAV14858.
XX
PT Enhancing immune response to an antigen
PT
XX
PS Example 12; Page 161; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 1211 BP; 217 A; 389 C; 394 G; 211 T; 0 other;
AA211303 Length: 1211 March 5, 2002 14:19 Type: N Check: 182 ..
Found using 'seq2-3' (pappu403.key)

23 ACGGTATGAGACTTCTTGACAGATGCTGGGCTTGGGACGCCGTTTGGCGTCGNGG
73 78

13 matches found in sequence:

```
83      CTGTCCGACACGCGATGATGCTGCTTGTGTGGCTTGAGAGGTTGGCGACCGCG
      93      98      132      137
143      GAGCATTCCTCCGCGCAGGTCGTCCGGTGGAGTACCTGATGCTGC
      ...
502      GGCACGGCCCTGTGCTGTGTGATGGCGGGTTCCGCCGCGCTGACCTGCGACCTGG
      552      557
562      CACCCGGAGCAATTATCTACGCGGGGCTGATGTCCGGTTCCTGACCCCTCCGAGGCG
622      TGTGGCCGTTCTGTATCAACATCTGATGGGTGGCGCGGCTTCAGAGCCGACGAC
      655      660      676      681
682      ATGTGGGCGAAGACCGAGGGGATCCCAACAGCGGTTGGACAGCGCAACGATCGATGTCG
      727      732
742      AACATCCGACCCGTGTGCGCCCAACAACCCGTATCTGGGTCTACTGGGGTAAAGGCCAG
      757      762
802      CCCACGAGCTCGGGGGGCGGACCTGCCCGCACGTTCTCTGAGGTTCTGACATCCGC
      820      825
862      ACCAAGAGACCTTCGCGGACACATCTCCGCCGGGGTGGCCACAGGTGTTCACAC
      889      894
922      TTCCCGGCCAACGGCAGCACAACGAGTGGGTTCGGAGCTTCGAGCGCATTCGAG
      972      977
982      CCTGACCTGACGGCGCACCTTCTGTGACGGTTGCACGAACGAAAGCCCGCGCATTCG
1042      GGGCGAGGTTTCGTCGTCGCGGGGCTACTGTGGCGACATAACGCAATCAACGGGATGG
      1055      1060
1102      TGCGTCATCAGAGACCGCGAGGGGTCAATTGCGCTACGACAGAGGTGGCGGACATCC
      1114      1119
1162      TTCCTGCC
      ...
9 matches found in sequence:
aaz11304 : M. vaccae antigen 85B encoding DNA.
(from "mycobacterin.seq")
TOIG of: aaz11304 check: 4039 from: 1 to: 485
ID      AAZ11304 standard; DNA; 485 BP.
XX
AC      AAZ11304;
XX
DT      25-OCT-1999 (first entry)
XX
DE      M. vaccae antigen 85B encoding DNA.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
```

```
OS      Mycobacterium vaccae.
XX
PN      W09932634-A2.
XX
PD      01-JUL-1999.
XX
PF      23-DEC-1998; 98MO-NZ00189.
XX
PR      04-DEC-1998; 98US-0205426.
PR      23-DEC-1997; 97US-0996624.
PR      23-DEC-1997; 97US-0997080.
PR      23-DEC-1997; 97US-0997362.
PR      11-JUN-1998; 98US-0095855.
PR      17-SEP-1998; 98US-0156181.
XX
PA      (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI      Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR      WPI; 1999-430163/36.
XX
DR      P-PSDB; AAY14859.
XX
PT      Enhancing immune response to an antigen
XX
PS      Example 12; Page 162; 243pp; English.
XX
CC      The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC      M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC      T cells and natural killer cells, to stimulate the production of
CC      cytokines, to enhance the expression of co-stimulatory molecules on
CC      dendritic cells and monocytes, and to enhance dendritic cell maturation
CC      and function. The proteins can be expressed by standard recombinant
CC      methodology. Pharmaceutical compositions comprising the proteins or
CC      nucleic acid sequences encoding the proteins can be used for the
CC      treatment, prevention, and detection of disorders including infectious
CC      diseases, immune disorders and cancer. In particular, the compounds and
CC      methods are used for treatment of diseases of the respiratory system,
CC      such as mycobacterial infections, asthma, allergies, tuberculosis,
CC      leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC      psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC      alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC      carcinoma and melanoma.
XX
SQ      Sequence 485 BP; 89 A; 165 C; 158 G; 73 T; 0 other;
      AAZ11304 Length: 485 March 5, 2002 14:19 Type: N Check: 4039 ..
Found using 'seq2-3' (pappu403.key)
      ...
44      ACTGGGTCTCGCGGTGATCATGCCCCGCGGGGCGCATCCAGCTTCTACAGCACTGGT
      94      99
104      ACAGCCCGCCTGCGGTAAAGCCGGCTCCAGACTTACAGTGGAGGACTTCTGAGACC
      150      155
164      AGGAGCTCGCGGCTACTCGCCGCCAACAGGGGGTGCAGCCGGAACGCAACGGGCGCG
224      TCGGTCTGTCATGAGCGCGTTTCGGGGCGCTGCGAGTGCATACACCCGCGAGCAGT
      249      254      261      266
284      TCAGTAGCGCGGGTTCGCTGTCGGGCTACTGAACCCGTCGAGGGGTGTGGCGGATGC
      297      302
344      TGATCAACATCTGATGGGTGACGCGGGGCTACAAAGCCACAGCATGTGGGGTCCAC
      385      390
```

404 CGAAGACCCGAGCGCGGTGAAGCGCAACCCGATGGTCATCGCACTGGCACTGG
418 423 433 438

464 TGGCCACAACACCCCTCTCTC

13 matches found in sequence:
aa211305 ; M. vaccae antigen 85C encoding DNA.
(from "mycobacteryn.seq")
TOIG of: aa211305 check: 742 from: 1 to: 1052

ID AA211305 standard; DNA; 1052 BP.

XX AA211305;

DT 25-OCT-1999 (first entry)

DE M. vaccae antigen 85C encoding DNA.

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.

XX Mycobacterium vaccae.

OS WO9332634-A2.

XX 01-JUL-1999.

XX 23-DEC-1998; 98WO-NZ00189.

XX 04-DEC-1998; 98US-0205426.

XX 23-DEC-1997; 97US-0996624.

XX 23-DEC-1997; 97US-0997080.

XX 23-DEC-1997; 97US-0997362.

XX 11-JUN-1998; 98US-0095855.

XX 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

DR WPI; 1999-430163/36.

XX P-PSDB; AAY14860.

PT Enhancing immune response to an antigen

PS Example 12; Page 162; 243pp; English.

XX The invention provides heat-killed Mycobacterium vaccae, or recombinant

CC M. vaccae proteins. The M. vaccae proteins may be employed to activate

CC T cells and natural killer cells, to stimulate the production of

CC cytokines, to enhance the expression of co-stimulatory molecules on

CC dendritic cells and monocytes, and to enhance dendritic cell maturation

CC and function. The proteins can be expressed by standard recombinant

CC methodology. Pharmaceutical compositions comprising the proteins or

CC nucleic acid sequences encoding the proteins can be used for the

CC treatment, prevention, and detection of disorders including infectious

CC diseases, immune disorders and cancer. In particular, the compounds and

CC methods are used for treatment of diseases of the respiratory system,

CC such as mycobacterial infections, asthma, allergies, tuberculosis,

CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as

CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,

CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell

CC carcinoma and melanoma.

XX Sequence 1052 BP; 185 A; 346 C; 342 G; 179 T; 0 other;

AA211305 Length: 1052 March 5, 2002 14:19 Type: N Check: 742
Found using 'seq2-3' (pappu403.key)

...

19 TTGTTTGCCGTTATGAAGTTTCACAGAGAAGTGCGGGGCTCCGCAAGGCGCGGATGCAC
69 74

79 CGGGTGGGGGCTTCCGATATATGCGCGCGTTGGCGCTCCCGGACTGATCGGCTTCCGGGG
85 90

139 GG

...

239 AGGGTGGCGGTACTCATGCGGTCTACCTGCTGACGCGTCTGCGTGGCCAGGACTTACA
289 294

299 ACGGCTGGGACATCAACACCCCTGCGTTCGAGTGTTCACGAGTCCGGCTTGTGACGA
354

359 TGCATGGCGGTGGCGGACAGTCCAGCTTCTACAGCGCATGGTACACGCCGTTCGGGGCA
359 396

419 ACGGCGAAGTACACCTTCAAGTGGAGACGTTCTCTGAACCCAGAGCTGCGGAGTGGC
447 452

479 TGGAGGCCAACCGCGGAGTGTGCGCACCGGACCGCTTCTGCGCTGTGATGGCGG
546 551

539 GCGGCGGCGGCTGACCTTACGCGATCCATCCACCCGAGCATTCATCTACGCGCTGCGG
546 551

599 TGTACAGGCTTCTGAACCCGTCGAGGCGTGTGGCCGATGCTGATCGGCGTGGATGA
599 656

659 ACGCCGAGGCGGCTTCAACGCCGAGAGCATGTGGGGCCCGTCTCGGACCGCGCTGGA
663 681

719 AGCGAAGCAGCCGATGTCATCATCACACGCTGTGGCCACAACACCGGATCTGGA
724 729

779 T

830 AGAAGCTGATGGCGGCGGACGTTCTCGAAGATTCACGTTGGCGACCAACACCGGCTTCC
880 885

890 GTGACAATACATCGCAGCGGCGGACCAACAGGTGTCTCAACTT

18 matches found in sequence:
aa211306 ; M. vaccae antigen GVC-7 encoding DNA.
(from "mycobacteryn.seq")

TOIG of: aa211306 check: 4195 from: 1 to: 795

ID AA211306 standard; DNA; 795 BP.

XX AA211306;

XX

```

DT 25-OCT-1999 (first entry)
XX
XX M. vaccae antigen GVC-7 encoding DNA.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX WO9332634-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-N200189.
XX
XX 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI WPI; 1999-430163/36.
XX
XX P-PSDB; AAY14861.
XX
XX Enhancing immune response to an antigen
XX
XX Example 13; Page 165; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
XX Sequence 795 BP; 132 A; 278 C; 272 G; 113 T; 0 other;
SQ
AAZ11306 Length: 795 March 5, 2002 14:19 Type: N Check: 4195
Found using 'seq2-3' (pappu403.key)
..
5 CGGGGTTGGCCATCTCTTGCGTCTGGGTCGGAGGCCATGTTGGGTACGATCCGG
|-----|
55 60
65 TACCGTCGGCGATGTGACCAACATCGAAGCAGACGAGCTAGAGAGCGGCGCTCG
|-----| |-----|
73 78 96 101 117 122
125 GCGCAGCAGCATTTGTGGCCGCCGCGGAGATGTCAGCGCGGAGCGGAGCCCGAGG
|-----| |-----|
165 170 175 180

```

```

185 AAGGCAACCAGSTCCGTTACACGCTCACCCTGCGCGGCTTACGATGTTCCAGCTTCT
|-----|
220 225
245 ATTCACACGAGAGCGCGCGGAGCATGAGGGGTTACAGCCGCGATATGCTTCGCA
|-----| |-----|
249 254 273 278 285
305 AGCGGAGAGAGTCAGCCTCGCCCGCGGTGTGCGGTCTTCGAACACAGATGCGCG
|-----|
365 ACCCGAAGTGGGCGATCCTTCAGGTACAGCAGCAGCAGCAGCGGTGGAGCGCGCCGGA
375 380
425 ACGCGCACTGCGAGCATCGCGTGCATGCGCAGAGAGTCTCAGCCAGCAGCAGCCCT
|-----| |-----|
439 444 475 480
485 ACAACGTGCGGTGCCAGCTCGGTAGTGTAGTCACCTCGCGAGAGTCCGCGCGAGCG
|-----|
545 CGGCGGCGAGCGGCTCGCGGTGACAGCAGCCCGAGCGCGTGGGTCCGCGGTACGCGCAG
545 577 582
605 GTAAAGATCGCTGGCGCCCGCGCGCCCTCGCGAGATCTGTCTCGGGGTAGACACAG
|-----|
611 616
665 CACGCGCTCTAATCTCCAGACCCCTTGCTGCGGTGGTGCACCGCGCCGCGGACACCGG
|-----|
668 673
726 GGGCGGATCACACGCTGTGCCCCCTCCGCTCCGCTCCGACGACGAGAAATCTCGCATG
|-----|
776 781
786 GCACCGGCGA
-----
5 matches found in sequence:
aaz11307; M. vaccae antigen GVC-8A (re-named GV-30) encoding DNA.
(from "Mycobactereng.seq")
TOIG of: aaz11307 check: 7397 from: 1 to: 300
ID AAZ11307 standard; DNA; 300 BP.
XX
XX AAZ11307;
AC
XX
XX 25-OCT-1999 (first entry)
DT
XX
XX M. vaccae antigen GVC-8A (re-named GV-30) encoding DNA.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX OS
XX PN
XX WO9332634-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-N200189.
XX
XX PF
XX 04-DEC-1998; 98US-0205426.
XX
XX PR 23-DEC-1997; 97US-0996624.

```



```

PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
XX P-PSDB: AA114862.
XX
XX Enhancing immune response to an antigen
XX
XX Example 13; Page 166; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 300 BP; 50 A; 96 C; 99 G; 55 T; 0 other:
XX
AAZ11307 Length: 300 March 5, 2002 14:19 Type: N Check: 7397 ..
Found using 'seq2-3' (pappu403.key)
..
..
50 CTTGCCCTTCCGAGAGCTGCGCGACGAGATTGCGATTCCTCGTATCTGCGCGCTGCTG
|-----|
100 105
103 108
110 CGGTATGGGGGACAGACGTCGACGGATCCGCCAGACCCGGCGCTCCGGACGGGC
170 GCGGTCTCGACATGTTCTGTCGTGCGGCGACACCTTCGAAAGGGTGGCGCTCAACACGA
|-----|
193 198 216 221
230 TCCAGTTCGCCGAGCTGTGCGCCGCCAGTTGTAGTGTTCGCCGAAATTGCATTCCAC
|-----|
235 240
290 G
...
7 matches found in sequence:
aaz11308; M. vaccae antigen GVC-8B (re-named GV-31) encoding DNA.
(from "mycobacterieng.seq")
FOIG of: aaz11308 check: 7652 from: 1 to: 563
ID AAZ11308 standard; DNA: 563 BP.
XX
XX AAZ11308;
XX
XX 25-OCT-1999 (first entry)
XX
XX M. vaccae antigen GVC-8B (re-named GV-31) encoding DNA.
DE

```

```

XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX W09932634-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-N200189.
XX
XX 04-DEC-1998; 98US-0205426.
XX 23-DEC-1997; 97US-0996624.
XX 23-DEC-1997; 97US-0997080.
XX 23-DEC-1997; 97US-0997362.
XX 11-JUN-1998; 98US-0095855.
XX 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
XX P-PSDB: AA114863.
XX
XX Enhancing immune response to an antigen
XX
XX Example 13; Page 166; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 563 BP; 85 A; 177 C; 200 G; 101 T; 0 other:
XX
AAZ11308 Length: 563 March 5, 2002 14:19 Type: N Check: 7652 ..
Found using 'seq2-3' (pappu403.key)
..
1 GATTCCTCGGCGGCTCAAGAGTCCGCGCGAGGTGATGTGACCGCTGACGGCTACGAG
|-----|
42 47
61 TTCAGTGGGCTTCCGAGGCGCTGTACACTTGCCTGGGACGAGTTCTGCGACTGTAT
|-----|
78 83
121 GTCGAGCTTGCCA
...
228 GTCACCGAGGCTGTGGAAGGCCCTGACCGGGCGGCGCCAGCGACGAGCTCTGGG
|-----|
278 283

```

```

288      AAATGTGAGTCACTGGTCTGGGAGCTGGCCACCGCCACCGGATACGGCTGATCA
      304      309
348      GGCTGCCGACACGACGATCCGCCACACCCAGAGTTGATCACGAGGTGCGCGGTTCCG
      364      369
408      CAGGATACGAGGTCTGGCCCGACCGCCAGCGGTGCTGCGCGGTTGTCGGCATGCACAC
      409      414
468      CGCGGTCTGGACGCCCATGTGCCCCGGGTGCGCGCGCTGCGCTTGACCGAGGCTG
      478      483
528      ATGAGG
...
-----
12 matches found in sequence:
aa211309 ; M. vaccae antigen GVC-8C (re-named GV-32) encoding DNA.
(from "Mycobacterieng.seq")
TOIG of: aa211309 check: 7613 from: 1 to: 434

ID      AA211309 standard; DNA: 434 BP.
XX
AC      AA211309:
XX
DT      25-OCT-1999 (first entry)
XX
DE      M. vaccae antigen GVC-8C (re-named GV-32) encoding DNA.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
OS      Mycobacterium vaccae.
XX
PN      W09932634-A2.
XX
PD      01-JUL-1999.
XX
PF      23-DEC-1998; 98WO-N200189.
XX
PR      04-DEC-1998; 98US-0205426.
PR      23-DEC-1997; 97US-0996624.
PR      23-DEC-1997; 97US-0997080.
PR      23-DEC-1997; 97US-0997362.
PR      11-JUN-1998; 98US-0098855.
PR      17-SEP-1998; 98US-0156181.
XX
PA      (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI      Prestige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR      WPI: 1999-430163/36.
XX
DR      P-PSDB; AAY14864.
XX
PT      Enhancing immune response to an antigen
XX
PS      Example 13; Page 166; 243pp; English.
XX
CC      The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC      M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC      T cells and natural killer cells, to stimulate the production of
CC      cytokines, to enhance the expression of co-stimulatory molecules on
CC      dendritic cells and monocytes, and to enhance dendritic cell maturation
CC      and function. The protein can be expressed by standard recombinant
CC      methodology. Pharmaceutical compositions comprising the proteins or

```

```

CC      nucleic acid sequences encoding the proteins can be used for the
CC      treatment, prevention, and detection of disorders including infectious
CC      diseases, immune disorders and cancer. In particular, the compounds and
CC      methods are used for treatment of diseases of the respiratory system,
CC      such as mycobacterial infections, asthma, allergies, tuberculosis,
CC      leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC      psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC      alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC      carcinoma and melanoma.
XX
50      Sequence 434 BP: 62 A; 139 C; 167 G; 66 T; 0 other;
AA211309 length: 434 March 5, 2002 14:19 Type: N Check: 7613
Found using 'seq2-3' (papp403.key)

1      GGGCCGGGCCCCGAGATGACCAAGTTCGAAGTCGTGCACCGGATGGGCTTTC
      31      36      45      50
61      GCCGACGCCGCCATGACGTCGCCGTGTCGAGGTGCGGGTCTGCTGGCTGGGACGCG
      76      81      85      90      107      112
121      ACGAAGCTGTGAACGACCGCGTGCATCACCCGATCGGGGTGACACACCGAC
181      TACCTCGGTGACACGATCGCCGAGATCGCCGCGGAGAAAGCCGGAATCATCACCCGCA
      196      201      205      210
241      GCCGACGACCTGTGTCGCCGACCGACCCGCTGCTGGCGCGAGATTCCGAGGCC
      244      249      268      273
301      ATGAGAGTGTGCTGCTGCCCGAGGC
...
335      ATGCGGCTGTAGCGCGGAGGATTGAGTGGCGGCTGTGGGCCCTGCAGGTGCGCATTCG
      385      390
395      GCGGACGCTGCTCGGTTGCAGGGGCTCGGTGGCTTAC
      426      431
-----
10 matches found in sequence:
aa211310 ; M. vaccae antigen GVC-8D (re-named GV-33) encoding DNA.
(from "Mycobacterieng.seq")
TOIG of: aa211310 check: 7163 from: 1 to: 438

ID      AA211310 standard; DNA: 438 BP.
XX
AC      AA211310:
XX
DT      25-OCT-1999 (first entry)
XX
DE      M. vaccae antigen GVC-8D (re-named GV-33) encoding DNA.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
OS      Mycobacterium vaccae.
XX
PN      W09932634-A2.
XX
PD      01-JUL-1999.

```

```

XX 23-DEC-1998: 98WO-NZ00189.
PF
XX
PR 04-DEC-1998: 98US-0205426.
PR 23-DEC-1997: 97US-0996624.
PR 23-DEC-1997: 97US-0997080.
PR 23-DEC-1997: 97US-0997362.
PR 11-JUN-1998: 98US-0095855.
PR 17-SEP-1998: 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI: 1999-430163/36.
XX P-PSDB: AAY14865.
XX
PT Enhancing immune response to an antigen
XX
PS Example 13; Page 166-167; 243pp: English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SO Sequence 438 BP; 61 A; 163 C; 146 G; 68 T; 0 other;
AAZ11310 Length: 438 March 5, 2002 14:19 Type: N Check: 7163
Found using 'seq2-3' (Papu403.key)
...
96 GGAGCAACAGATGTCGCCCTCAGCGGCGTGGACCGCAGCGCATGCAGACGTCCCGG
156 TGGCGCTGAGGCGCTTCCCGCCGCTGAGCGCGGTGTGATGCCAACGACACTACGACC
166 171 194 199 200 205
216 ACCTGCACATCGACACCATGTGCGGTTGGCGGCACACCCAGCGGCCCCCTGTGTGTC
233 238
276 CGTTGGGCAATCGCGCGACACACTCTCGCAAGTGGGCGTCCCGAGGCGCGGATGTGAGT
308 313 326 331
336 TGGACTGGCAGCAGAGCCACCGCATCGACGACCTGACGCTGTGACACCCCGCCCGGC
362 367 375
396 ACTTCTCGGCGCGGTGTTCTCCCGGCACTCGACGCTGTGGC
427 432
-----
1 match found in sequence:

```

```

aaz11313 : Cloned insert sequence from M. vaccae genomic DNA.
(from "Mycobacterieng.seq")
TOIG of: aaz11313 Check: 1266 from: 1 to: 313
ID AAZ11313 standard; DNA: 313 BP.
XX
XX AAZ11313;
AC
AC AAZ11313;
XX
DT 25-OCT-1999 (first entry)
XX
DE Cloned insert sequence from M. vaccae genomic DNA.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX W09932634-A2.
PN
XX
XX 01-JUL-1999.
PD
XX
XX 23-DEC-1998: 98WO-NZ00189.
PF
XX
XX 04-DEC-1998: 98US-0205426.
XX 23-DEC-1997: 97US-0996624.
XX 23-DEC-1997: 97US-0997080.
XX 23-DEC-1997: 97US-0997362.
XX 11-JUN-1998: 98US-0095855.
XX 17-SEP-1998: 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
XX WPI: 1999-430163/36.
XX
PT Enhancing immune response to an antigen
XX
PS Example 12; Page 170; 243pp: English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SO Sequence 313 BP; 64 A; 105 C; 96 G; 48 T; 0 other;
AAZ11313 Length: 313 March 5, 2002 14:19 Type: N Check: 1266
Found using 'seq2-3' (Papu403.key)
...
68 ATGTGCGAGGAGCCCGTGGCGGTGGCGGCGCTCGAACATCCGAGTGTGACACGCTGTAC
118 123
128 GGCTGCACATGTGGGCGCGAGCTCAATCCGCAAGTAACCTGTGTGAC

```

1 match found in sequence:

aa211318 ; M. vaccae antigen GVC-7 specific primer.
(from "mycobacterieng.seq")
TOIG of: aa211318 check: 8171 from: 1 to: 33

ID AA211318 standard; DNA; 33 BP.

AC AA211318;

DE 25-OCT-1999 (first entry)

DE M. vaccae antigen GVC-7 specific primer.

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; PCR primer; ss.

OS Synthetic.

OS Mycobacterium vaccae.

PN W09932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-N200189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0996624.

PR 23-DEC-1997; 97US-0997080.

PR 23-DEC-1997; 97US-0997362.

PR 11-JUN-1998; 98US-0095855.

PR 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

WPI; 1999-430163/36.

XX Enhancing immune response to an antigen

XX Example 13; Page 173; 243pp; English.

XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

SO Sequence 33 BP; 11 A; 9 C; 12 G; 1 T; 0 other;

AA211318 Length: 33 March 5, 2002 14:19 Type: N Check: 8171
Found using 'seq-3' (pappu403.key)

|-----|

1 GAGAGACTGAGACGCCCGAGGAGGACACAG
13 18

16 matches found in sequence:

aa211321 ; Partial DNA sequence of M. vaccae antigen GV-33.
(from "mycobacterieng.seq")
TOIG of: aa211321 check: 4320 from: 1 to: 825

ID AA211321 standard; DNA; 825 BP.

AC AA211321;

DE 25-OCT-1999 (first entry)

DE Partial DNA sequence of M. vaccae antigen GV-33.

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.

OS Mycobacterium vaccae.

PN W09932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-N200189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0996624.

PR 23-DEC-1997; 97US-0997080.

PR 23-DEC-1997; 97US-0997362.

PR 11-JUN-1998; 98US-0095855.

PR 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

WPI; 1999-430163/36.

XX P-PSDB; AAY14874.

XX Enhancing immune response to an antigen

XX Example 13; Page 174; 243pp; English.

XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

SO Sequence 825 BP; 123 A; 297 C; 271 G; 134 T; 0 other;

AA211321 Length: 825 March 5, 2002 14:19 Type: N Check: 4320
Found using 'seq-3' (pappu403.key)

...

```

98      GGAGCAACAGATGTTCGCCCTCAAGGGCGGTGAGCCGACGCATCAGACGATGCCCGG      |-----|
158      TGCCGCTGAGGGCGCTTCCCGCGGTGAGCGCGGTGTGATCAGCCACGACACTAGACACC      |-----|
      168      173
218      ACCTCGACATCGACACCACCATGTGCGGTGGCGCACACCCAGCGGGCCCCGTTGTGTGTC      |-----|
      235      240
278      CGTTGGGCAATCGGCGCACACCTGCGCAATGGGGCGTCCCGAGGCGCGATGTCAGT      |-----|
      310      315      328      333
338      TGGACTGGCAGCAAGCCACACCGCATAGACGCTGAGCTGTGCACCCCGCCCGGCGC      |-----|
      364      369      377
398      ACTTTCGGAGCGGTTGTTCTCCCGGACTCGACGCTGTGGCGGTGTGGTGCACCG      |-----|
      429      434      438      443
458      GCTGTGCGCACAGGGGCTTCTGGTGGCGCACACCGGATACAGAAAGCTTCCCGCAGA      |-----|
      471      476      484      489
518      TCGCGCAGCAGTACGGGTCCGTTGATCTGACCCCTGCTGCCGATCGGGGCTTACCATTCC      |-----|
      520      525
622      CATCTGACCTGACGAGGTGAGCAACAGCCTGATGTCCTCCATCCACTGGCGACATTC      |-----|
      672      677
682      CGCCTGCCCCCGATCCGTGTCGAGACCCCGGACCGCTGACCGCTGCCGACGCC      |-----|
      716      721      736      741
742      GAGCGGGTACGCTGACCGTGCAGATTCGCCGTGAGCGGTTGAGCCCGGAGTGCAGCTTC      |-----|
      795      800
802      GACCCGTGTGGCGGTTTGAAACC
-----
8 matches found in sequence:
aa211322 : M. vaccae GroEL-homologue clone GV-27 DNA sequence.
(from "mycobacteriing.seq")
TOIG of: aa211322 check: 1254 from: 1 to: 337

ID      AA211322 standard; DNA: 337 BP.
XX
AC      AA211322:
XX
DT      25-OCT-1999 (first entry)
XX
DE      M. vaccae GroEL-homologue clone GV-27 DNA sequence.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
OS      Mycobacterium vaccae.
XX
PN      WO932634-A2.

```

```

XX      01-JUL-1999.
PD
XX
PF      23-DEC-1998;      98MO-NZ00189.
XX
XX      04-DEC-1998;      98US-0205426.
PR      23-DEC-1997;      97US-0996624.
PR      23-DEC-1997;      97US-0997080.
PR      23-DEC-1997;      97US-0997362.
PR      11-JUN-1998;      98US-0095855.
PR      17-SEP-1998;      98US-0156181.
XX
XX      (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX      Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
XX      WPI; 1999-430163/36.
XX
XX      P-PSDB; AAY14876.
XX
XX      Enhancing immune response to an antigen
XX
XX      Example 14; Page 175; 243pp; English.
XX
XX      The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX      M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX      T cells and natural killer cells, to stimulate the production of
XX      cytokines, to enhance the expression of co-stimulatory molecules on
XX      dendritic cells and monocytes, and to enhance dendritic cell maturation
XX      and function. The proteins can be expressed by standard recombinant
XX      methodology. Pharmaceutical compositions comprising the proteins or
XX      nucleic acid sequences encoding the proteins can be used for the
XX      treatment, prevention, and detection of disorders including infectious
XX      diseases, immune disorders and cancer. In particular, the compounds and
XX      methods are used for treatment of diseases of the respiratory system,
XX      such as mycobacterial infections, asthma, allergies, tuberculosis,
XX      leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX      psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX      alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX      carcinoma and melanoma.
XX
XX      Sequence 337 BP; 60 A; 112 C; 108 G; 57 T; 0 other;
XX
XX      AA211322 length: 337 March 5, 2002 14:19 Type: N Check: 1254
XX      Found using 'seq2-3' (pappu403.key)
-----
50      TGCTCCCGTGTGAGAAAGTCATCCAGCGCGCAAGCGCTGTGATCATCGCGGAG      |-----|
      100      105      109
110      ACCTCGAGGGCGAGGCCCTGTCCACGCTGTGTCAACAAGATCCGGCGACCTTCAAGT      |-----|
      114
170      CCGTCGCCCTCAAGGCTCCGGGCTTGGGTACCCGCCCAAGGCGATCTCAGACATGC      |-----|
      172      177      210      215
230      CCATCTCACCGGTGTGTCAGGTGATCAGCGAAGAGTCGGGTGTCCTGAGAGACGCGG      |-----|
      250      255
290      ACCTTCGCTGTGTGGCCAGCGCCGCAAGGTCGTCATCACCAGACAGACA      |-----|
      294      319      324      327

```

11 matches found in sequence:
aa211323 : M. vaccae 65 kd heat shock protein partial DNA sequence.
(from "mycobacteriing.seq")

TOIG of: aaz11323 check: 6469 from: 1 to: 300

ID AAZ11323 standard; DNA; 300 BP.
 AC AAZ11323;
 XX
 XX
 DT 25-OCT-1999 (first entry)
 DE
 XX M. vaccae 65 kd heat shock protein partial DNA sequence.
 XX
 KM Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
 KM dendritic cell maturation; infectious disease; immune disorder; cancer;
 KM respiratory system; mycobacterial infection; allergy; tuberculosis;
 KM leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KM dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KM squamous cell carcinoma; melanoma; ss.
 XX
 OS Mycobacterium vaccae.
 XX
 PN W09932634-A2.
 PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-NZ00189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX
 DR WPI: 1999-430163/36.
 DR P-PSDB: AAY14877.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Example 14; Page 176; 243pp; English.
 XX
 CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
 CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.
 CC
 SQ Sequence 300 BP; 60 A; 104 C; 98 G; 38 T; 0 other;
 AAZ11323 Length: 300 March 5, 2002 14:19 Type: N Check: 6469 ..
 Found using 'seq2-3' (pappu403.key)

1 CCCTACGAGAAAGATCGCGCTGAGTGTCTCAAGAGATCGCAAGACCGACGAGTC
 16 21 37 42 52 57 60
 |-----|
 61 GCGGCGACGGGACCAACACCGCCACCGTGTCTGCTCAGGCTGTGTTGCGGAAGGCTG

64 69
 121 CGCAACGTCGCGAGCCGCGCAACCCGCTCGGCTCAAGCGTGCATCGACAGAGCTGTCTC
 124 129 136 141
 181 GAGGCTGTACACCAGTCGCTGTGAGTGGCCAGAGAGGTGAGACCAAGACAGATT
 195 200
 241 TCTGCACACGGGCGATCTCGCGGAGACACCCAGATCGGCGAGCTCATCGCGAGGCC
 252 257 265 270 289 294
 301

 1 match found in sequence:
 aaz11324; Mycobacterial 16S rRNA specific oligo AD10.
 (from "mycobacterieng.seq")
 TOIG of: aaz11324 check: 6401 from: 1 to: 43

ID AAZ11324 standard; DNA; 43 BP.
 AC AAZ11324;
 XX
 DT 25-OCT-1999 (first entry)
 XX
 DE Mycobacterium 16S rRNA specific oligo AD10.
 XX
 KM Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
 KM dendritic cell maturation; infectious disease; immune disorder; cancer;
 KM respiratory system; mycobacterial infection; allergy; tuberculosis;
 KM leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KM dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KM squamous cell carcinoma; melanoma; PCR primer; ss.
 XX
 OS Synthetic.
 OS Mycobacterium vaccae.
 OS
 PN W09932634-A2.
 PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-NZ00189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX
 DR WPI: 1999-430163/36.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Example 15; Page 176; 243pp; English.
 XX
 CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
 CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and

CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

XX Sequence 43 BP; 13 A; 7 C; 18 G; 5 T; 0 other;

AAZ11324 Length: 43 March 5, 2002 14:19 Type: N Check: 6401 ..
Found using 'seq2-3' (pappu403.key)

1 ACTGACGCTGAGGAGCGAAGCGTGGGAGCGAAGAGATTAG
4 9

5 matches found in sequence:
aaz11329; M. vaccae band 12B21 (pota gene homologue) nucleotide sequence.
(from "mycobacterieng.seq")
TOIG of: aaz11329 check: 9855 from: 1 to: 238

ID AAZ11329 standard; DNA; 238 BP.

XX AAZ11329;

DT 25-OCT-1999 (first entry)

DE M. vaccae band 12B21 (pota gene homologue) nucleotide sequence.

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.

XX Mycobacterium vaccae.

PN WO9932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0996624.

PR 23-DEC-1997; 97US-0997080.

PR 23-DEC-1997; 97US-0997362.

PR 11-JUN-1998; 98US-0095855.

PR 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

XX WPI: 1999-430163/36.

XX P-PsDB; AAY14878.

XX Enhancing immune response to an antigen

XX Example 15; Page 177-178; 243pp; English.

CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and

CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

XX Sequence 238 BP; 52 A; 77 C; 71 G; 38 T; 0 other;

AAZ11329 Length: 238 March 5, 2002 14:19 Type: N Check: 9855 ..
Found using 'seq2-3' (pappu403.key)

1 CTCGATGTAACCGCTCGGAGCGCTCGACCTGACCTGCGCCACGTCATGCAGTTCGAGCTC
18 23

61 AGCCGATCCAGCGGAGGTGCGGATCAGCTTCATCTACGTGACCCAGCAGCAAGAG
120

121 GCGCTCAGCATGAGTACGCCATCGCGGTGATGACGCGCAGCATCGAGATCGGC
125 154 159 163 168

181 AGCCGACCGAGATCTACGACGCTCCCGCAGCGGTGTGTCGCGCAGCTTCATCGAAT
220 225

26 matches found in sequence:
aaz11330; M. vaccae pota gene homologue nucleotide sequence.
(from "mycobacterieng.seq")
TOIG of: aaz11330 check: 6563 from: 1 to: 1518

ID AAZ11330 standard; DNA; 1518 BP.

XX AAZ11330;

DT 25-OCT-1999 (first entry)

DE M. vaccae pota gene homologue nucleotide sequence.

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.

XX Mycobacterium vaccae.

PN WO9932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0996624.

PR 23-DEC-1997; 97US-0997080.

PR 23-DEC-1997; 97US-0997362.

PR 11-JUN-1998; 98US-0095855.

PR 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

XX WPI: 1999-430163/36.

XX P-PsDB; AAY14879.

XX Enhancing immune response to an antigen

PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI WPI: 1999-430163/36.
XX
XX Enhancing immune response to an antigen
PT
XX
XX Example 15; Page 180; 243pp; English.
PS
XX The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 323 BP; 76 A; 101 C; 97 G; 49 T; 0 other;
AA211333 Length: 323 March 5, 2002 14:19 Type: N Check: 1920 ..
Found using 'seq2-3' (pappu403.key)
1 GTGCACTACAAAGACGCTTCAACGACGAGCAGCACTGGTGGCCAAAGGTCAAGAGACCG
22 27
61 TTGTGCGCAGAGGAGACATAGGCGCGGCGGCTGTGATCCGCCAGGTTCATGCGCGCG
82 87
121 CGCGTCAAGGGCCTGGG
...
257 TGGGTCTCGCCTACACAAAGGACGACCGGACGATATCCGACCATGAGACCTCT
307 312
317 GGGATCC

19 matches found in sequence:
aa211334; *M. vaccae* pold gene homologue nucleotide sequence.
(from "mycobacterium.seq")
TOIG of: aa211334 check: 711 from: 1 to: 1341
ID AA211334 standard; DNA; 1341 BP.
XX
XX AA211334;
AC
XX
XX
DT 25-OCT-1999 (first entry)
XX
XX
DE *M. vaccae* pold gene homologue nucleotide sequence.
XX
XX *Mycobacterium vaccae* protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;

KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
XX *Mycobacterium vaccae*.
OS
XX
XX W09932634-A2.
XX
XX
XX 01-JUL-1999.
PD
XX
XX 23-DEC-1998; 98WO-NZ00189.
PF
XX
XX 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI WPI: 1999-430163/36.
XX
XX P-PSDB; AAY14880.
DR
XX
XX Enhancing immune response to an antigen
PT
XX
XX Example 15; Page 180-181; 243pp; English.
PS
XX
XX The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
XX
SQ Sequence 1341 BP; 264 A; 463 C; 408 G; 206 T; 0 other;
AA211334 Length: 1341 March 5, 2002 14:19 Type: N Check: 711 ..
Found using 'seq2-3' (pappu403.key)
...
49 TGACATGATCCCGACCTGTGGCCGGAATGACCGGACCGGACCTGGGTGCGCGCTT
99 104
109 CATGGCGGTGGCGCGCGCGCGCGCGCGCGCTGACCTGCTGCTGCTGCGCGCGC
119 124
169 GTGCGGTCGACAGTGGGACCTCGAGCACCACGCTACAGACAGGCGGCCCCCAGCGG
227
229 CGCCGTGCGGTCTCCAACTGGCGGCTATATGGCGGAGCGTTTCATGCGCAGCTTCCA
232 280 285
289 GACCGCTCGGCGCATCGGTGACCTCAAGAGACTTCAACGACGACGAGCGTGT
329 334


```

489 GAGAGATGACCCGGCTGTGTGAGAGCCCTCCAGACCTTACTGTGCGGCGCGC
      539 544
549 ACCTTCGCGAAGGTATCTCCGATGGGATGCCGGGGTGTGCGCGGAGCATGCTG
      575 580
609 GTGTTATCCCGCGCTCGGACCTTCATCAACGCCGACTATCTCGGCATACCCAGACC
      627 632 639 644
669 ACCATGATCGGCACAGCTGATCCAGAGCAGTTCTGTCGTCAGAGACTATCCGCGCGC
      705 710
729 GCGCGCTGAGTCTGGGGCTGATGTTGCTGATCTGATCGCGCTGCTCTTACACACGG
      788
789 GCGCTGGGTTCGGAGAGATCTGTATGACCAACCAGCAGCGCGCCACTGCGCACCCGCGC
      793 827 832
849 CCCAGCAGGATCC

-----
10 matches found in sequence:
aaz1138 ; M. vaccae potC gene homologue (GV-26) nucleotide sequence.
(from "mycobacterieng.seq")
TOIG of: aaz1138 check: 5163 from: 1 to: 277

ID AA211338 standard; DNA; 277 BP.
XX
AC AA211338;
XX
DT 25-OCT-1999 (first entry)
XX
DE M. vaccae potC gene homologue (GV-26) nucleotide sequence.
XX
KW Mycobacterium vaccae proteini; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO9932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI: 1999-430163/36.
XX
P-PSDB: AAY14882.
XX
PT Enhancing immune response to an antigen
XX
PS Example 15; Page 183-184; 243pp; English.
XX

```

```

CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 277 BP; 44 A; 95 C; 89 G; 49 T; 0 other;
AA211338 Length: 277 March 5, 2002 14:19 Type: N Check: 5163
Found using 'seq2-3' (pappu403.key)

...
17 CCCGTACGCCGCTAGGCAAACTCATGGGTTTCGCTCAAGACTTCAGAGGCGCAGCATCTCG
      67 72
77 GTGCGGTGGGATCAAGAGGCGGCGCTGAGAAAGCTTCCCGCGCTGCGACAGCCGTGCTC
      84 89 94 99 118 123 124 129
137 GTCTCGTCAAGTATCATGATGAGGCCAAGTCTCTCCGCGCGCGCCAGATGCCCCG
      178 183
197 CCCGCCAGGCGCGCTGCGCGGATCCCGTGAGACGTCGCTGCGTACATGACACA
      211 216 232 237 246 251
257 AGATGTGCTGTGCGGCATCG

-----
10 matches found in sequence:
aaz11339 ; Nucleotide sequence of M. vaccae partial clone GV-4P.
(from "mycobacterieng.seq")
TOIG of: aaz11339 check: 4027 from: 1 to: 485

ID AA211339 standard; DNA; 485 BP.
XX
AC AA211339;
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae partial clone GV-4P.
XX
KW Mycobacterium vaccae proteini; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO9932634-A2.
XX
PD 01-JUL-1999.
XX

```

PF 23-DEC-1998; 98WO-N200189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX
 DR WPI: 1999-430163/36.
 DR P-PSDB; AAY14873.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Example 12; Page 186; 243pp; English.
 XX
 CC The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
 CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.
 XX
 S0 Sequence 485 BP; 88 A; 166 C; 158 G; 73 T; 0 other;

AAZ11339 Length: 485 March 5, 2002 14:19 Type: N Check: 4027 ..
 Found using 'seq2-3' (pappu403.key)

```

44  ACTGGGTCTCGGGGTGATCATGCCCGCTCGGGCGGAGTCACCTTCTACAGCGACTGCT
    |-----|
    94  99

104  ACAGCCCGGCGCTGCGTAAGCGGCTGCCAGACCTTACAGTGGAGACGTTCTCTGACCC
    |-----|
    150 155

164  AGAGAGCTCGCGGCTTACCTGCGCCCAACAAGGGGGTGCAGCCGACCGACGCGGCGCG
    |-----|
    249 254 261 266

224  TCGGCTGTGTCATGCGCGGTTTCGGCGGCTGACGCTGCGGAGCTTACACCGCCAGAGT
    |-----|
    297 302

284  TCCAGTACCGCGGCTGTCTCGGGCTTACCTGAACCGGTCGAGGGGTGTGGCCGATGC
    |-----|
    385 390

344  TGATCAACATCTCGATGGGTGACGCGGGGCTTACAAGGCCCATGTGGGGTGCACA
    |-----|
    418 423 433 438

404  CCGAGAGACCGGAGCAGCGCTTGAAGCGACGACCGGATGTCAACATCGGCAAGCTGG
    |-----|
    463

```

464 TCGCCACACACACCCCTCTC
 468

9 matches found in sequence:
 aaz11340 : Nucleotide sequence of *M. vaccae* antigen GVC-13.
 (from "mycobacterieng.seq")
 TORG of: aaz11340 check: 2024 from: 1 to: 501

ID AAZ11340 standard; DNA; 501 BP.
 XX
 AC AAZ11340;
 XX
 DT 25-OCT-1999 (first entry)
 XX
 DE Nucleotide sequence of *M. vaccae* antigen GVC-13.
 XX
 KW *Mycobacterium vaccae* protein; antigen; T cell activation; cytokine;
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KW squamous cell carcinoma; melanoma; ss.
 XX
 OS *Mycobacterium vaccae*.
 XX
 FN W09932634-A2.
 XX
 PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-N200189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX
 DR WPI: 1999-430163/36.
 DR P-PSDB; AAY14888.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Example 10; Page 186; 243pp; English.
 XX
 CC The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
 CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.
 XX
 S0 Sequence 501 BP; 69 A; 178 C; 174 G; 78 T; 2 other;

AAZ11340 Length: 501 March 5, 2002 14:19 Type: N Check: 2024 ..
 Found using 'seq2-3' (pappu403.key)

```

14      GTGCGCGCAGTGCGGCTTGCGCTCCGTCGCTCGCGCGCGTGCGTGGCGGTGAGAG
      |-----|
      87 92 95      107 112      123 128
74      GCACCGCACTGGCGCGCGGAGCTGGAGCGGCGCTACACGGTGTGACGTTGCGCTT
      |-----|
      87 92 95      107 112      123 128
134     CCGACAACTCGGCGACGAGTGTGGCGCGCGCGCCGACAGCAACCCG
      |-----|
199     TTCAGCAGCTCCTGTGTGGGACACTGCTGGCCACCGCGCTCCGACGCGCCCGCGCTGC
      |-----|
      249 254
259     AACCGCAGATTCGCGACGCCCGCGGCGCTACACCTGGGAGCGGACGAGTGTGTTCAAC
      |-----|
      264 269
319     TACAACCTGGCAGTGGAGTGTCTCCGCGCGCGACGTCGCCGCGAGTACGCCCGCGCG
      |-----|
      346 351      352 357
379     CGTTCGCTGTGTCTTACGCCCGCGACCGC
      |-----|

4 matches found in sequence:
aa211342 : Nucleotide sequence of M. vaccae antigen GVs-9.
(from "mycobacterieng.seq")
TOIG of: aa211342 check: 8101 from: 1 to: 503

ID      AA211342 standard; DNA; 503 BP.
XX
AC      AA211342;
XX
DT      25-OCT-1999 (first entry)
XX
DE      Nucleotide sequence of M. vaccae antigen GVs-9.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
OS      Mycobacterium vaccae.
XX
PN      WO9932634-A2.
PD
XX      01-JUL-1999.
XX
PF      23-DEC-1998; 98WO-N200189.
XX
XX      04-DEC-1998; 98US-0205426.
XX      23-DEC-1997; 97US-0996624.
XX      23-DEC-1997; 97US-0997080.
XX      23-DEC-1997; 97US-0997362.
XX      11-JUN-1998; 98US-0095855.
XX      17-SEP-1998; 98US-0156181.
XX
PA      (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI      Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX

```

```

DR      WPI; 1999-430163/36.
DR      P-PSDB; AAY14890.
XX
PT      Enhancing immune response to an antigen
XX
PS      Example 11; Page 187-188; 243pp; English.
XX
XX      The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC      M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC      T cells and natural killer cells, to stimulate the production of
CC      cytokines, to enhance the expression of co-stimulatory molecules on
CC      dendritic cells and monocytes, and to enhance dendritic cell maturation
CC      and function. The proteins can be expressed by standard recombinant
CC      methodology. Pharmaceutical compositions comprising the proteins or
CC      nucleic acid sequences encoding the proteins can be used for the
CC      treatment, prevention, and detection of disorders including infectious
CC      diseases, immune disorders and cancer. In particular, the compounds and
CC      methods are used for treatment of diseases of the respiratory system,
CC      such as mycobacterial infections, asthma, allergies, tuberculosis,
CC      leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC      psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC      alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC      carcinoma and melanoma.
XX
SQ      Sequence 503 BP; 69 A; 162 C; 183 G; 88 T; 1 other;
XX
      AA211342 length: 503 March 5, 2002 14:19 Type: N Check: 8101 ..
      Found using 'seq2-3' (pappu403.key)

...
215     TGGGCACCTACGCGGTGTAACCTACCGCGCAGATTCCGACTTCACAATCGCGCCGATG
      |-----|
      265 270
275     GCGCGCGCGACGCATCGGGCGCGGTGCACTGATGCGCGACAACCTGCCCGGACACCAAG
      |-----|
335     CTTGTCTCTGGCGGCATGTGCGCANGGCGCGCGGTATGACCTGATCACTCGATCCG
      |-----|
      359 364      365 370
395     CGACCGCTGGCGCGGTTCAACCCCGCATGCCGCCCGCGCTGCCCGCACCGATGGCC
      |-----|
      437 442
455     GCGGTGTGTGTTTCGAAATCCGTTGCGCGACATCCG
      |-----|

50 matches found in sequence:
aa211343 : Nucleotide sequence of M. vaccae antigen GV-27.
(from "mycobacterieng.seq")
TOIG of: aa211343 check: 1714 from: 1 to: 1569

ID      AA211343 standard; DNA; 1569 BP.
XX
AC      AA211343;
XX
DT      25-OCT-1999 (first entry)
XX
DE      Nucleotide sequence of M. vaccae antigen GV-27.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
OS      Mycobacterium vaccae.
XX

```

```
XX XX
PN MO9932634-A2.
XX XX
PD 01-JUL-1999.
XX XX
PF 23-DEC-1998; 98WO-NZ00189.
XX XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX XX
DR WPI: 1999-430163/36.
DR P-PSDB; AAY14891.
XX XX
PT Enhancing immune response to an antigen
XX XX
PS Example 14; Page 188-189; 243pp; English.
XX XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX XX
SQ Sequence 1569 BP; 292 A; 519 C; 531 G; 227 T; 0 other;
AAZ11343 Length: 1569 March 5, 2002 14:19 Type: N Check: 1714
Found using 'seq2-3' (pappu403.key)
```

```
308 TCGCAACGTCGACCGCGCAACCCGCTTCGACCTCAAGCTGCATCGAAGAGGCTG
313 318 325 330
368 TCGAGGCTGTCAACCCAGTCGCTGAAAGTGGCCCAAGAGGTCCAGACCAAGAGCAGA
384 389
428 TTTCTGCCACCGCGCGGATTTCCGCGGACACCCAGATCGCGAGCTCATCGCCGAGG
441 446 454 459
488 CCATGACAAAGTGGCGCAACGAGGCTGTATCACCCTGAGAGATC
...
563 CCGAGGATATGCGCTTCGACAAAGGCTACATCTCGGGTTACTTCTGACCCGCGGAC
613 618
623 GCGAGAGCCGCTCTCGAGAGATCCCTACATCTCTGTGTCAGTCCAAAGTG
625
692 TGCTCCGCTGTGGAGAAAGTCATCCAGGCGGACCGCTGTGATCATCGCGGAG
742 747 751
752 ACGTCAGAGGCGAGGCGCTGTCCACGCTGTGTCAACAGATCCGCGCACCTTCAAGT
756
812 CCGTCGCGTCGAAGCTCCGGGCTTCCGTGACCCCGCAAGCGCATGCTGCAGACATAG
814 819 852 857
872 CCATCTCACCGGTGTGTCAGCTGTCAAGCAAGAGTGGGCTGTCCCTGGAGACCGCGG
892 897 931
932 ACGTCTGCTGTGGGCGAGCGCCGCAAGGTGTCTACCAAGAGACGACCAACATCG
936 961 966 988
992 TCGAGGGCTGGGCGGATTCGATGCCATCGCGCGGTGGTTCAGATCCGCGCGAGA
993 1003 1008 1018 1023
1052 TCGAAGAACAGCACTCCGACTACGACCGCGAAGACTGTCAGAGACCGCTGCCAAGCTGG
1060 1065 1094 1099
1112 CCGCGGCTGTGCGGTATCAAGCGCGAGCTGCCACCGAGTGGAGCTCAAGAGCGCA
1172 AGCACCGCATAGAGACGCGCTCCGCAACGCGAAGGCTGCCGTGAAGAGGCGATCGTG
1186 1191 1225 1230 1228
1232 CCGGTGGCGCGGTGCTGTCTGTAGTCGAGTCCCTGCGCTGGAGACCTCGGCTGACGG
1233 1273 1278
1292 GCGACGAGCGCCACCGGTGCCAATCTGCTCGCGGTGGGCTGTGGCTCCGCTCAAGCAGA
1296 1315 1320 1326 1331 1351
```

```

1352 1356 1378 1383 1381 1386
-----|-----|-----|
TCGCTTCACAGCGCGGCTGGAGCCCGGCTGCTGGCCGCAAGAGTGTCCACCTGCCCC

1412 CGGCTACAGCGCTCAGCCGCGGATGACGAGACCTGCTCAGAGCGCGGCTGCG
1426 1431 1465 1470 1468
-----|-----|-----|
-|
1472 CCGACCCCGGTGAAGTGCACCCGCTCGGCTGCGAGAACGCGGCTCATCGCGCTCTGT
1473 1497 1502 1512 1517

1532 TCCTCACACCGAGCGCGTGTGCTGCGCAGACACCGCGAG
1549 1554 1552 1557
-----|-----|-----|

20 matches found in sequence:
aa211344 ; Nucleotide sequence of M. vaccae antigen GV-27A.
(from "Mycobacterium.seq")
TOIG of: aa211344 check: 6231 from: 1 to: 647

ID AA211344 standard; DNA: 647 BP.
XX AA211344:
XX
XX
XX 25-OCT-1999 (first entry)
XX
XX Nucleotide sequence of M. vaccae antigen GV-27A.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX OS
XX
XX PN WO9932634-A2.
XX
XX PD 01-JUL-1999.
XX
XX PF 23-DEC-1998; 98WO-NZ00189.
XX
XX PR 04-DEC-1998; 98US-0205426.
XX PR 23-DEC-1997; 97US-0996624.
XX PR 23-DEC-1997; 97US-0997382.
XX PR 23-DEC-1997; 97US-0997382.
XX PR 11-JUN-1998; 98US-0095855.
XX PR 17-SEP-1998; 98US-0156181.
XX
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI, 1999-430163/36.
XX DR P-PSDB; AAY14892.
XX
XX PT Enhancing immune response to an antigen
XX
XX PS Example 14; Page 190-191; 243pp; English.
XX
XX CC The invention provides heat-killed Mycobacterium vaccae, or recombinant

```

```

CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
XX SQ Sequence 647 BP: 136 A; 204 C; 215 G; 92 T; 0 other:
XX
XX AA211344 Length: 647 March 5, 2002 14:19 Type: N Check: 6231 ..
Found using 'seq2-3' (pappu403.key)

...
8 AGACAATTGCGTATGACGAAGAGCGCCGCGTGGCTCGAGCGGGGCTCAACGCCCTCG
58 63
68 CAGACGCCGTAAGGTACGTTGGGCGCGAAGGTGCGCATGCTGGAAGAAGT
70 75 84 89 106 111
128 GGGCGCCGCCGACGATCACCAAGATGTGTGTCATCGCCAGAGAGATGAGCTGAGG
130 135 148 153 163 168
188 ACCGATGAGAGATCGGCGCTGAGCTGTCAAGAAGTGCACCAAGACGACGAGC
205 210 226 231 241 246
248 TCGGGGCGAGCGGACCAACCGCACCGCTGCTCGCTCAGGCTGTGTTGCGAAGGCC
249 253 258
308 TGGCAACGTCGACAGCGGCGCCAAACCGCTCGGCTCAAGCGTGCATCGAGAAAGCTG
313 318 325 330
368 TCGAGGCTGTACCACTGCTGCTGAATGCGCCAAAGAGTGCAGACCAAGAGCAGA
384 389
428 TTTCTGCGACCGGCGATTTTCGCGCGGACACCGAGATCGGGAGCTCATCGCGAGG
441 446 454 459 478 483
488 CCATGACAAAGTGGACAGAGGGGTCACTACCGTCAAGAGATC
...
563 CCGAGGATATGCGCTTGCAGACAGGCTACATCTCGGGTTACTTCTGACGACCGCGAGC
613 618
623 GCGAAGAACCGTCTGAGAGATCC
625
-----
30 matches found in sequence:

```

aa11345 ; Nucleotide sequence of M. vaccae antigen GV-27B
(from "mycobactereng.seq")
TOIG of: aa11345 check: 3042 from: 1 to: 927

ID AA211345 standard; DNA; 927 BP.

AC AAZ11345;

DT 25-OCT-1999 (first entry)

Nucleotide sequence of *M. vaccae* antigen GV-27B.

KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.

05 *Mycobacterium vaccae*.

PN W09932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0997080.

PR 11-JUN-1998; 98US-0095855.

XX XX

[illegible][illegible]

DR P-PSDB; AAY14893.

PT Enhancing immune res

PS Example 14; Page 22

CC The invention provides heat-killed Myco

CC T cells and natural killer cells, to

CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

SQ Sequence 927 BP; 157 A; 317 C; 317 G; 136 T; 0 other;

```
AA211345 Length: 927 March 5, 2002 14:19 Type: N Check: 3042
Found using 'seq2-3' (pappu403.key)
```

50. TGCTCCGCTGTGAGAAGTTCATCCAGGCCGCAAGCCGCTGCTATCATCGCCGAG
100 105 109

110 -----
ACGTCGAGGGCGAGCCCTGTCCACGCTGGTGTCACACAGATCCGCGGCACCTTCAAGT

170 CCGTCGCCITCAAGGCTCCGGGCTTCGGTGAACCGCCGCAAGGCGGATCTGCAGACATGG
172 177 210 215

230 CCAATCTCACCGGTGGTCAGGTGTCAGCGAAGAGTTCGGCTGTCCCTGGAGACCGCCG
250 255 289

290
294
319 324
322 327
346

350 TCAGAGGCTCGGCGCATTCGATGCCATCGCCGCGGGTGGCTCAGATCCGCGCCGAGA
351 361 366 376 381

410 TCGAGACAGCGACTCCGACTACGACCGGAGAGCTGCAGAGCGCTTGCCCAAGCTGG
418 423 452 457

470 CCGCGGTGTTGCGGTGATCAAGGCCGAGCTGCCACCGAGGTGGAGCTCAAGGAGCGCA

530 AGCACCGCATTCGAGGACGCCGTCGCCAAGCGGAGGCGATCGTCG
544 549
583 586

590 CCGGTGGCGCGGTGGCTTGTGTCAGATCGGCTCTCCGCTGGACGACCTCGGCTGACGG
631 636 649

650
654
673 678 684 689 709

-----|-----|-----|
GCGAGGAGGCCGCGTCCACATCGTCCGCGTGGCGTGTGGCTCCGCTCAAGCAGA

710
714
736
741
739
744

770
CGGGTCACGGCCTCAACGGCGGACCGGTGATGATACGAGGACCTGCTCAAGGCGCCGCGTGG
784 789
826

830
831

- |

CGACCCGATGAAGTCAACCGCTGCGGGCTGCAGAACCGCGCGTCCATTCGCGGCTGT

855 860 870 875

890 TCCTCACCAGAGCCGTCGTGGCCGACAGCCGAG
907 912
910 915

7 matches found in sequence:
 azal1346 ; Nucleotide sequence of M. vaccae 3S-PCR band 12B28
 (from "mycobacterium.seq")
 ToIG of: azal1346 check: 88 from: 1 to: 162


```

ID AA211346 standard; DNA; 162 BP.
XX
AC AA211346;
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae 3S-PCR band 12B28.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO9932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI; 1999-430163/36.
XX
PT Enhancing immune response to an antigen
XX
PS Example 15; Page 193; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 162 BP; 25 A; 53 C; 51 G; 33 T; 0 other;
AA211346 Length: 162 March 5, 2002 14:19 Type: N Check: 88
Found using 'seq2-3' (pappu403.key)

1 CTCGTACAGCGACGAGATCTCCGACGCGTACGCGTTCGCGGCGTTCATCAGAGGC
9 14 25 30 49 54
1-----|-----|-----|
61 TTGCGGCTGCTGTTGTTGTTGCTGCTGCGGCGTCAACGCGACCCCTCATCAGAGGC
64 69 78 83 100 105
1-----|-----|-----|

```

```

121 GCACGAGAACAGACTGGCGGACGCGATTCCGTC
...
-----
31 matches found in sequence:
AA211347; Nucleotide sequence of M. vaccae antigen GV-38A.
(from "Mycobacterium.seq")
TOIG of: aa211347 check: 5325 from: 1 to: 1366

ID AA211347 standard; DNA; 1366 BP.
XX
AC AA211347;
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae antigen GV-38A.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO9932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI; 1999-430163/36.
XX
DR P-PSDB; AAT14894.
XX
PT Enhancing immune response to an antigen
XX
PS Example 15; Page 193; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 1366 BP; 210 A; 447 C; 464 G; 243 T; 2 other;
AA211347 Length: 1366 March 5, 2002 14:19 Type: N Check: 5325
Found using 'seq2-3' (pappu403.key)

```

```

...
143 CTGGCTACCTACATCTGCGCTGGCGCGTTGCTCTCTGCTGTACAGCGCGATGGAG
      |-----|
      193 198
203 ATCTCCGACGACGCGCAGCTGGGTACGGTTGGCTCCGACCCCTGTTCCGGCTGTGTTG
      |-----|
      209 214 212 217 233 238 248 253 262
263 ACGTTGGGTGTCTCCGGGCTCAAGCCACCTCATCCAGGGCGCACGAAAGACAGCTGG
      |-----|
      267 284 289
323 CGCAGGGGATTCCTGATCTTCTCTGACGCTCGCGCTTCGCGCTGATGCGGCTGCGT
      |-----|
      350 355
383 ATCACCGTGAATCAGGCGCTATGTGTGGGGCGCGAAGCTGGGGGCGCTTTCACCGCACTG
      |-----|
      443 448 455 460
443 GCGCTACCTTCATCTTCTTGGCGCTCTGCGAATTGCTGCGTCAAGATCATCTCG
      |-----|
503 GGTCGTGCTGCTGTTTGAAGCAACGTTCCGGCTCGGCGACTGGATCCACGCTCCGACC
      |-----|
      539 544
563 GCGCGGGCGCGCGCTCCGCCACAGCGCGCGT
      |-----|
606 ACTGGCGTCAACATATGACAGACCGCGCAACTGTGTATGATGCCACGCGCGAAC
      |-----|
      656 661
666 TCGCCGCGCGCTGCTTACCAATTACAGCGCGCGCGTGGAGAGACCGGCGTGAACCGTGC
      |-----|
      676 681
726 TCACCACTTCAACGCGCGGACACCGCCGATGATGTCTGCGAGATGCTTCGCGCTGCG
      |-----|
      727 737 742 775 780
786 CGGCGTCTGCGCGCAACTGCGACCGACGAGACGATCCGCCACGCTTATCTCGGTCGCG
      |-----|
      787 792 790 795 821 826
846 CCGAATACGAGAAGTGTATCCGTTGCACACACCGCGGCTGGAGTCTGTCAGAGACA
      |-----|
      887 892
906 CGTACCTGGGATGGTGTGTGTACCGCGCGCGCGGAGAACTTCCCTTAAACGGGCTGCG
      |-----|
      959 964 962
- |-----|
966 CCGACGANTTTCAGACCGCGGAGAGGATGCGTCCGCGCATGCGGCTGTGTCACACAC
      |-----|
      967 992 997 1015 1020
1026 TCGCGTTGGCAGCGCAACGACGAGAGATGCGCGACGTGTGCTGTGTCGCTTACG
      |-----|
      1037 1042 1055 1060

```

```

1086 GCACGGGGAGCCCTCCAGACGCCGGTCAAGTACCGACCGGATGAGTTTCATCTGAG
      |-----|
      1095 1100
1146 ACGGACGGGTGAGTCTGTCCGTGATCATCAGACGCGATGATCCCGCGCGGTGC
      |-----|
      1181 1186
1206 TCGAGCGTGGAGCTTCTCTGGGGCGAGACCGCTGACCGCGGGAACCGGTACTCGGACCG
      |-----|
      1214 1219 1258 1263
1266 CGCAGCGCGTGGAGAGATCAACGCTGCTGGAGATGCGCCGTGACGAGATCGAGCGCTGG
      |-----|
      1317 1322
1326 TGACACCGAAGCCGATCTGCTGCAAGTATGATGGGCGCGTGC
      |-----|
16 matches found in sequence:
aaz11348 ; 5' nucleotide sequence of M. vaccae antigen GV-38B.
(from "mycobacterieng.seq")
TOIG of: aaz11348 check: 210 from: 1 to: 898
ID AAZ11348 standard; DNA; 898 BP.
XX
XX AC
XX AAZ11348;
XX
XX DT
XX 25-OCT-1999 (first entry)
XX
XX DE
XX 5' nucleotide sequence of M. vaccae antigen GV-38B.
XX
XX MYcobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; ss.
XX OS
XX Mycobacterium vaccae.
XX
XX PN
XX WO9932634-A2.
XX
XX PD
XX 01-JUL-1999.
XX
XX PF
XX 23-DEC-1998; 98WO-N200189.
XX
XX PR
XX 04-DEC-1998; 98US-0205426.
XX 23-DEC-1997; 97US-0996624.
XX 23-DEC-1997; 97US-0997080.
XX 23-DEC-1997; 97US-0997362.
XX 11-JUN-1998; 98US-0095855.
XX 17-SEP-1998; 98US-0156181.
XX
XX PA
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX PI
XX Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
XX DR
XX MPI; 1999-430163/36.
XX
XX DR
XX P-PSDB; AAY14895.
XX
XX PT
XX Enhancing immune response to an antigen
XX
XX PS
XX
XX Example 15; Page 195; 243pp; English.
XX
XX CC
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX CC methodology. Pharmaceutical compositions comprising the proteins or
XX CC nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious

```



```

206 TCACCACTTCAACGCCGGGACACCCCGCATGATGTCGAGATGCTGCTGCGTGC
    207      217 222      255 260
      |-----|
      |-----|
266 CGGCGTGGCTGCCCCGAATGCGCAGCAGGAGACATGCGCCTTATCTCGTGGCG
    267      272      301 306
      270 275
      |-----|
326 CCGAATACGAGAAATCGATCCGTTGCACACACCCCGGTGGTACGAGACGA
    367      372
      |-----|
      |-----|
386 CGTACCTGGATGGTGTGTGTACCGCCGGCGGCAAGACTTCCGCTTAACGCGCTGC
    438      443
      441
      |-----|
      |-----|
446 CGACGATTCGACACGCCGGAGACGATGCCCTCGGCCATCGGGCTGTGGCTCACACTG
    447      452      470 475      493 498
      446
      |-----|
      |-----|
506 CGCTTGGCAGACGACAGACGAGATGCGGACGTGTGTGCTGTGCTTACGCGC
    515      520      533 538
      |-----|
      |-----|
566 AACGGGAACGCTCCAGCAGCGGGTCAGGTACCGACCGGATGAGCTTCATCGTAGAC
    573      578
      |-----|
      |-----|
626 GGCAGGGTGAGTGTCTGCTGATCGATCAGGACGCGACGTGATCCCGCGGGTGTCTC
    659      664
      |-----|
      |-----|
686 GAGCGTGGGACATTCCTCGGGGACAGACACGCTGACGGGACCGGTACGCGGACCGCG
    692      697      736 741
      |-----|
      |-----|
746 CACGCGCTGGAGGAAGTACACCGTGTGAGATGAGCCCGTGACGAGATCAGCGCTGTG
    795      800
      |-----|
      |-----|
806 CACCGAAAGCCGATCCTGCTGCTGACGTGATCGGGGCGCTG
    795      800
      |-----|
      |-----|
14 matches found in sequence:
    (from "Mycobacterieng.seq")
    TOIG of: aaz11357 check: 2437 from: 1 to: 742

```

```

XX 23-DEC-1998: 98MO-NZ00189.
XX PF
XX 04-DEC-1998: 98US-0205426.
XX PR
XX 23-DEC-1997: 97US-0996624.
XX PR
XX 23-DEC-1997: 97US-0997080.
XX PR
XX 23-DEC-1997: 97US-0997362.
XX PR
XX 11-JUN-1998: 98US-0095855.
XX PR
XX 17-SEP-1998: 98US-0156181.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Prestidige RL, Skinner MA, Tan P, Vlaser ES, Watson J;
XX PI
XX XX
XX WPI: 1999-430163/36.
XX DR
XX P-PSDB; AA114898.
XX
XX Enhancing immune response to an antigen
XX
XX Example 15; Page 200; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 742 BP; 138 A; 243 C; 228 G; 133 T; 0 other:
XX
    AAZ11357 Length: 742 March 5, 2002 14:19 Type: N Check: 2437 ..
    Found using 'seq2-3' (pappu403.key)
    ...
100 GACCTGAAGAATCGATGTGATTTACTCGCGGACAGCAGCTGCCAGGAGCGATCGGC
    150 155
    |-----|
160 GCGTTGACGAGCGTTTCCGTCAGCTGGCGATCGGACGATACCGGGGACAGCGCG
    166 171      187 192      200      216
    |-----|
    |-----|
220 TCATTGGCGCGTTACTACGACCGGACGTTGCCCAACACCAACCTTCGACGACGGAAC
    221      243 248      265 270
    |-----|
    |-----|
280 CGCGTGCAGCTCCGCGCGCTCATCCGAATCCAAACCCCGACGCGCTATCTGACGCGCTC
    286 291      320 325      333 338
    |-----|
    |-----|
340 TATACCCCGCGTTTCAGAACCTGGAGAGAGCGATCGCGTTGCGAGAGCGCGGACGCG
    369 374      382 387
    |-----|
    |-----|
400 AGCGCTGTGTGCGCGCGCAATGCCAGATTCAACGAGTTCTTCGCGGAGATGTGCA
    400 405
    |-----|
    |-----|
...
494 AGGCAACGTGTGTACTCCGCTACAAAGGGCGGATCTCGGAGCAAAACATCGTCAACG
    494

```

554 GCCCTATCGCAACGGGAACTGTGGAAGCTACGAGAGCGGT

544 549

1 match found in sequence:

aa211358 ; 5' nucleotide sequence of M. vaccae antigen GV-42.
(from "mycobacterieng.seq")
TOIG of: aa211358 check: 4939 from: 1 to: 45

ID AA211358 standard; DNA; 45 BP.

AC AA211358;

DT 25-OCT-1999 (first entry)

DE 5' nucleotide sequence of M. vaccae antigen GV-42.

KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;

KW dendritic cell maturation; infectious disease; immune disorder; cancer;

KW respiratory system; mycobacterial infection; allergy; tuberculosis;

KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;

KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;

KW squamous cell carcinoma; melanoma; ss.

OS Mycobacterium vaccae.

PN WO9932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0996624.

PR 23-DEC-1997; 97US-0997080.

PR 23-DEC-1997; 97US-0997362.

PR 11-JUN-1998; 98US-0095855.

PR 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

PI WPI: 1999-430163/36.

DR P-PSDB; AAY14899.

PT Enhancing immune response to an antigen

PS Example 16; Page 202; 243pp; English.

XX The invention provides heat-killed Mycobacterium vaccae, or recombinant

CC M. vaccae proteins. The M. vaccae proteins may be employed to activate

CC T cells and natural killer cells, to stimulate the production of

CC cytokines, to enhance the expression of co-stimulatory molecules on

CC dendritic cells and monocytes, and to enhance dendritic cell maturation

CC and function. The proteins can be expressed by standard recombinant

CC methodology. Pharmaceutical compositions comprising the proteins or

CC nucleic acid sequences encoding the proteins can be used for the

CC treatment, prevention, and detection of disorders including infectious

CC diseases, immune disorders and cancer. In particular, the compounds and

CC methods are used for treatment of diseases of the respiratory system,

CC such as mycobacterial infections, asthma, allergies, tuberculosis,

CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as

CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,

CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell

CC carcinoma and melanoma.

SO Sequence 45 BP; 7 A; 13 C; 15 G; 9 T; 1 other;

AA211358 Length: 45 March 5, 2002 14:19 Type: N Check: 4939 ..

Found using 'seq2-3' (pappu403.key)

1 ATGAGCGAATGCGCCGNCCTGCGGGGTGTGCATGTGCATC
10 15

5 matches found in sequence:

aa211359 ; Nucleotide sequence of M. vaccae antigen GV-43 middle fragment.
(from "mycobacterieng.seq")
TOIG of: aa211359 check: 6957 from: 1 to: 340

ID AA211359 standard; DNA; 340 BP.

AC AA211359;

DT 25-OCT-1999 (first entry)

DE Nucleotide sequence of M. vaccae antigen GV-43 middle fragment.

KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;

KW dendritic cell maturation; infectious disease; immune disorder; cancer;

KW respiratory system; mycobacterial infection; allergy; tuberculosis;

KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;

KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;

KW squamous cell carcinoma; melanoma; ss.

OS Mycobacterium vaccae.

PN WO9932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.

PR 04-DEC-1998; 98US-0205426.

PR 23-DEC-1997; 97US-0996624.

PR 23-DEC-1997; 97US-0997080.

PR 23-DEC-1997; 97US-0997362.

PR 11-JUN-1998; 98US-0095855.

PR 17-SEP-1998; 98US-0156181.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

PI WPI: 1999-430163/36.

DR P-PSDB; AAY14900.

PT Enhancing immune response to an antigen

PS Example 16; Page 202; 243pp; English.

XX The invention provides heat-killed Mycobacterium vaccae, or recombinant

CC M. vaccae proteins. The M. vaccae proteins may be employed to activate

CC T cells and natural killer cells, to stimulate the production of

CC cytokines, to enhance the expression of co-stimulatory molecules on

CC dendritic cells and monocytes, and to enhance dendritic cell maturation

CC and function. The proteins can be expressed by standard recombinant

CC methodology. Pharmaceutical compositions comprising the proteins or

CC nucleic acid sequences encoding the proteins can be used for the

CC treatment, prevention, and detection of disorders including infectious

CC diseases, immune disorders and cancer. In particular, the compounds and

CC methods are used for treatment of diseases of the respiratory system,

CC such as mycobacterial infections, asthma, allergies, tuberculosis,

CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as

CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,

CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell

CC carcinoma and melanoma.

SO Sequence 340 BP; 50 A; 127 C; 121 G; 40 T; 2 other;

AA211359 Length: 340 March 5, 2002 14:19 Type: N Check: 6957 ..

XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI WPI; 1999-430163/36.
DR P-PSDB; AAY14902.
XX
XX Enhancing immune response to an antigen
PT
PS
PS Claim 3; Page 203-204; 243pp; English.
XX
XX The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
-CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis, alopecia
CC areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

SQ Sequence 273 BP; 43 A; 105 C; 91 G; 34 T; 0 other;
 AA211361 Length: 273 March 5, 2002 14:19 Type: N Check: 6219
 Found using 'seq2-3' (Pappu403.key)

14 AGGGGGTTCGGCCGATCGAGGCCCGGGTGGCCGACAGCGGATACAGCAACGCCCGGG

74 CCAAGGGCTACTTCCCCGCTGAGCTTCACCGTGGCCGGCATGACCAAGCGTCCGATCG
|----|
103 108

134 TGACCGCAACGTACACCGCGGCGGCCGAGCGGGCGCCGTGGCCACCCAGCCGCTGACC
142 147 156 171 189

194 TCATCGCCCGGAGCCCGACCGATGGCAGCTGTCCAAGCAGTCCGCACTGGCCCTGA
194 201

254 TGTCCGGTCAATGCCGCA
265 270

7 matches found in sequence:
aaz11362 ; Nucleotide sequence of M. vaccae antigen GVC-22B.
(from "Nucleotide sequence of M. vaccae antigen GVC-22B")

TOIG of: aaz11362 check: 5978 from: 1 to: 554

ID AA211362 standard; DNA; 554 BP.

AC AAZ11362;

DT 25-OCT-1

DE Nucleotide sequence of *M. vaccae* antigen GVC-22B.

KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;

KW respiratory system; mycobacterial infection; allergy; tuberculosis;

dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;

KW squamous cell carcinoma; ss.

OS Mycobacterium vaccae.

PN W09932634-A2.

PD 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.

| | | |
|----|--------------|---------------|
| PR | 04-DEC-1998; | 98US-0205426. |
| PR | 03-DEC-1997; | 07US-0005534 |

| | | |
|----|--------------|---------------|
| PR | 23-DEC-1997; | 97US-0997080. |
| PR | 23-DEC-1997; | 97US-0997080. |

PR 11-JUN-1998; 98US-0095855.
DE 17-SEP-1998; 98US-0155191

[illegible]

XX

XX 1000 420253 230
EFT 1000 420253 230

DR P-PSDB; AAY14903.

PT Enhancing immune response to an antigen

PS Claim 3; Page 204; 243pp; English.

CC The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.

SQ Sequence 554 BP; 112 A; 187 C; 178 G; 77 T; 0 other;

AAZ11362 Length: 554 March 5, 2002 14:19 Type: N Check: 5978
Found using 'seq2-3' (pappu403.key)

1 GATGTCACCCCGAGATGTACGTCGACCGAGACGCCGTCGSCACACAGATTAC
22 27 37 42

61 GTTGAGCACTTCAGATCTGGTTACCTTGA

•

142 AACATGAAATTCACCTGGAATGACCGTGCGCGCAAGCCGGCGCCCTGGCCGGCGTGGGG

202 CGGCATGTCGTTCGGCGGCGTGGCCGGGCAACCGTGGCGGGCACAGATGGCGGGCGCCC 255 260

— : —

276 281

— — — — —


```
382 AGCCGCGCCGAGCGGAGCAACCTGCGGGCTACTTCAACG
...
11 matches found in sequence:
aa211363 : Nucleotide sequence of M. vaccae antigen GV-1/83.
(from "mycobacterng.seq")
FOIG of: aa211363 check: 9501 from: 1 to: 808

ID AA211363 standard; DNA; 808 BP.
AC AA211363;
XX
XX
XX 25-OCT-1999 (first entry)
DE Nucleotide sequence of M. vaccae antigen GV-1/83.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
OS
XX WO992634-A2.
PN
XX 01-JUL-1999.
PD
XX 23-DEC-1998; 98WO-NZ00189.
PF
XX 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095655.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI WPI; 1999-430163/36.
XX P-PSDB: AAY14904.
DR
XX Enhancing immune response to an antigen
XX
XX Claim 3; Page 205; 243pp; English.
PS
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 808 BP; 140 A; 266 C; 279 G; 122 T; 1 other;
XX
XX AA211363 Length: 808 March 5, 2002 14:19 Type: N Check: 9501
Found using 'seq2-3' (pappu403.key)
```

```
1 CCAAGTGTGACGGGNGTGTGACGGTATCCGACCAATCCACAGACGCCGACACTGTG
27 32 44 49 52
61 GAATCACCCGTTGCAATTCAATGATGGGGCAACGGTGTCCGT
...
111 GGATTCAGGAAATGATGACAACTCGCCGGAAGTCAAGCCGAGTGGGGGAATGCTGCGG
161 166
171 TGCCATCTCTCGGTGGCGCCGCAATGTTCCAGTGAAGACGGTGGGAG
...
233 AGCAGCAGCGGCTCTCCGGATGAGTCCGCGCAGCAGAGATGACACAGTGTGCGG
283 288
293 GCCCTTTCGGCCGACCCCTGCGGCCAACTGATCGGCTCCGGCTGCG
...
362 GAAGTCCCGGGTGGTGGCGGGAGTGGCAGCCGATCCGATGAGTGGCGGCTGGAAC
412 417
422 AACCCGATGCTGACAGCGTGTCCGACGAGGCTCTCCGGCCAGCTCAATCCGAGTCAAT
436 441 448 453
482 CTCGTGACACCCCTCGACGGCGGTGAGTTTCAACCGTGTGCGCGCAGCAGCAGCGCTTC
530 535
542 GCCAAGATGATCGGCGCACGCTGGAGACCCCTCAAGACGACTTC
...
626 GGCCAGGCGCGCCCGCATCATGAGTGTGCGCGACATGTACGTTGGAGGGGCGCGCTTC
676 681
686 ACGGTCTCCGGGATGGCCGACCAAGTCAAGGTCAAGCAGCGCGGTGTGTGCGGTGGG
719 724
746 GTGCAGACCGCCCAACGCGAGCGGTGTATCT
...
1 match found in sequence:
aa211366 : M. vaccae DNA fragment obtained by PCR amplification.
(from "mycobacterng.seq")
FOIG of: aa211366 check: 8916 from: 1 to: 102

ID AA211366 standard; DNA; 102 BP.
XX
XX AA211366;
AC
XX 25-OCT-1999 (first entry)
DE M. vaccae DNA fragment obtained by PCR amplification.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
```



```
287      GGTACGCTCAACCCGAACGTCAATCTGTCGACACGTTCAACGGCGCCAGTTCCCGTG
      302 307
347      TTGGCGCCGCAACATGACGCTTCGGCCAGATGATGATCCGGGCCACGCTGAGACCCCTCAAG
      362 367
407      ACCGATTCCGA
...
455      GGCCAGCGCCGCCGATCAGGTGCTGGCGGAGCATGTGACGGTGTGAGGGGGCGCCGGTC
      505 510
515      ACGGTTCGGGATGGCGCCGACGCTCAAGTCAAGTCACGCGCTGTGTGTGGGTGGG
      548 553
575      GTGCAGACCGCCCAACGCGACGCTGTATCT
...
-----
11 matches found in sequence:
aaz11368 ; Extended DNA sequence for M. vaccae antigen GVs-9.
(from "Mycobacterium.seq")
TOIG of: aaz11368 check: 4381 from: 1 to: 1125

ID      AAZ11368 standard; DNA; 1125 BP.
XX
AC      AAZ11368;
XX
DT      25-OCT-1999 (first entry)
XX
DE      Extended DNA sequence for M. vaccae antigen GVs-9.
XX
KW      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW      dendritic cell maturation; infectious disease; immune disorder; cancer;
KW      respiratory system; mycobacterial infection; allergy; tuberculosis;
KW      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW      squamous cell carcinoma; melanoma; ss.
XX
OS      Mycobacterium vaccae.
XX
PN      W09932634-A2.
XX
PD      01-JUL-1999.
XX
PF      23-DEC-1998; 98WO-NZ00189.
XX
PR      04-DEC-1998; 98US-0205426.
PR      23-DEC-1997; 97US-0996624.
PR      23-DEC-1997; 97US-0997080.
PR      23-DEC-1997; 97US-0997362.
PR      11-JUN-1998; 98US-0095855.
PR      17-SEP-1998; 98US-0156181.
XX
PA      (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI      Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR      WPI: 1999-430163/36.
DR      P-PSDB; AAY14906.
XX
PT      Enhancing immune response to an antigen
XX
PS      Claim 3; Page 208; 243pp; English.
XX
CC      The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC      M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC      T cells and natural killer cells, to stimulate the production of
```

```
CC      cytokines, to enhance the expression of co-stimulatory molecules on
CC      dendritic cells and monocytes, and to enhance dendritic cell maturation
CC      and function. The proteins can be expressed by standard recombinant
CC      methodology. Pharmaceutical compositions comprising the proteins or
CC      nucleic acid sequences encoding the proteins can be used for the
CC      treatment, prevention, and detection of disorders including infectious
CC      diseases, immune disorders and cancer. In particular, the compounds and
CC      methods are used for treatment of diseases of the respiratory system,
CC      such as mycobacterial infections, asthma, allergies, tuberculosis,
CC      leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC      psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC      alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC      carcinoma and melanoma.
XX
SQ      Sequence 1125 BP; 159 A; 389 C; 406 G; 170 T; 1 other;
      AAZ11368 Length: 1125 March 5, 2002 14:19 Type: N Check: 4381 ..
      Found using 'seq2-3' (pappu403.key)
```

```
215      TGGGCACCTACGCCGTGTACTACCCGGCAGGATTGGAATTTCACAATGCGCGCCCATG
      265 270
275      GGCGCGCGCCGACGATCGGGGGGAGTGACAGTGTGCGCCGACAACTGCCCGACACCAAG
      359 364
335      CTGTGCTGTGGCGGATGTGCGANGCGCGCGCGCTCATGCACTGATCACCGTGTGATCCG
      359 364
395      CGACCGCTGGCGCGTTACCCGCCACCCGATGCGCGCCCGCGTGCAGCACCGTGGCC
      437 442
455      GCGGTGTGTGTTTCGGAATTCGTTGCGGACATCCG
...
510      CGCAGATGAGGGGACCTACGAGGCGCGAATGATGATCTGTGTGCGCTCGACGATCCGT
      560 565
570      TCTGCTCGCCCGGCTTAACCTGCGGCGCACTTGCCTACGCCGA
...
618      ACGGCATGTGTGAGGAAGCCGCCAATCTTCGCCCGCTTGAAACGGGGCCAGAGCGTGAGC
      668 673
678      TGCCCGAGGGGCGCTTACCTGACACTGTTTCGTCGCCGGGGGAGGTAAAGCTGGAGGAGC
      685 690
738      CCGGACCGCTGCGCGAAGGCGAGCGACGACGTGGCTTACCGGATCGGGCGCGCGGTGA
      739 755
798      CCGCACCGCGCC
...
1017      GCGACGCGCGTTTCACACAGCCCGGAGGCGCTGTGCGCGAGGGTGGAGCGCTGCG
      1067 1072
1077      GTGTGGCGCGGACCGCGCGCGGCTGAGCGCGTGGGCGCCCGG
...
```

7 matches found in sequence:

aa211369 ; Nucleotide sequence of M. vaccae antigen GV-35.
(from "mycobacterieng.seq")
TOIG of: aa211369 check: 7565 from: 1 to: 666

ID AA211369 standard; DNA; 666 BP.
AC AA211369;
XX
DT 25-OCT-1999 (first entry)
DE
XX Nucleotide sequence of M. vaccae antigen GV-35.
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
PN W09932634-A2.
XX
PD 01-JUL-1999.
XX
FE 23-DEC-1998; 98WO-N200189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI; 1999-430163/36.
DR P-PSDB; AAY14907.
XX
XX
PT Enhancing immune response to an antigen
PS
XX Claim 3; Page 211; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 666 BP; 108 A; 224 C; 209 G; 125 T; 0 other;
AA211369 Length: 666 March 5, 2002 14:19 Type: N Check: 7565
Found using 'seq2-3' (pappu403.key)

7 GCAATCATTCGGAGATGCTACAAATCCGCGCGCGATATGTCGATCCATCCGCGCTT

57 62
67 TGTTCGCCCGACGTGGACACATCTCATGAGGTCTGGACAAATGAGCTGAGCCTGGGACATC
127 CACGGCCAGGGCGCCGAGACGACTGACCAATTCAGCAGTGGAGACCTTCTCTCAACGCGTC
143 148 181 186
187 TTCCCGTTGGACACCGACCGGTTGACCGGAGTGGTTCCACTCGGACGACGACTTAC
247 GTCTGTGGCCGGTGAAGTGTCCGACGAGTTTCAGAGGCGACGCTGAGCTGAGCTACAGGTG
307 GGCCTTCCGTGTGCTGTGGCGTGGGATTCACACTTACGCTACACCCCGCAACATCAGC
318 323
367 TACGACG
572 CTGCGCGTGTGTGTGCTGCTGCGTCCGCTTCGCCCGCTATCTCTGTCACCGCGACGCG
622 627
628

9 matches found in sequence:

aa211370 ; Nucleotide sequence of M. vaccae antigen GV-5P.
(from "mycobacterieng.seq")
TOIG of: aa211370 check: 8598 from: 1 to: 480

ID AA211370 standard; DNA; 480 BP.
XX
AC AA211370;
XX
DT 25-OCT-1999 (first entry)
DE
XX Nucleotide sequence of M. vaccae antigen GV-5P.
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
PN W09932634-A2.
XX
PD 01-JUL-1999.
XX
FE 23-DEC-1998; 98WO-N200189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;

XX WPI: 1999-430163/36.
DR P-PSDB: AAY14908.
XX
PT Enhancing immune response to an antigen
PS Claim 3; Page 212; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 480 BP; 90 A; 156 C; 156 G; 78 T; 0 other;
AAZ11370 Length: 480 March 5, 2002 14:19 Type: N Check: 8598 ..
Found using 'seq2-3' (pappu403.key)
..
7 TGGGACATCAACACCCCTCGCTTCAGAGTGTCTACAGATCCGGCTTGTGAGATCATG |-----|
57 62
67 CCGGTGGCGGAGACAGTCTTACAGCAGCTGTACACGAGCGGTCTGGGGCAACGGG |-----|
94 99
127 CAGAACTACACACTCAAGAGTGGAGAGCGTTCAGACCCAGAGAGCTGCCAGCTGGAGAG |-----|
150 155
187 GCCAACCGCGGAGTGTGCGCCACCGGAAAGCGCTTCGCTGTGATGCGGGGAGAC |-----|
247 GCGGCGCTGACCTACGCGGATCATCAACCGCAGATTCAATACGCTTCGCTGTCA |-----|
249 254 297 302
307 GGCTTCTGAACCCGTCGAGGGCTGTGCGCCATGCTGATCGGGCTGGGATGAACGAC |-----|
354 359 366
367 GCAGCGGCTTCAACGCCGAGCATGTGGGCCCGCTCTCGAGCCGCGGTGAAGCGC |-----|
379 384
427 AACGACCCGATGTCAACATCAACAGCTGTGTGGCAACAACACCGGATCTGG |-----|
427 432
52 matches found in sequence:
aaz11371; Nucleotide sequence of M. vaccae antigen GV-27.
(from "mycobacterieng.seq")
TOIG of: aaz11371 check: 7168 from: 1 to: 1626
ID AAZ11371 standard; DNA; 1626 BP.
XX
AC AAZ11371;
XX
DT 25-OCT-1999 (first entry)

XX
DE Nucleotide sequence of M. vaccae antigen GV-27.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN WO932634-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-NZ00189.
XX
XX 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
DR P-PSDB: AAY14909.
XX
PT Enhancing immune response to an antigen
PS Claim 3; Page 212-213; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 1626 BP; 302 A; 537 C; 552 G; 235 T; 0 other;
AAZ11371 Length: 1626 March 5, 2002 14:19 Type: N Check: 7168 ..
Found using 'seq2-3' (pappu403.key)
..
8 AGACAATTGGTATGACGAAGAGCCCGCGCTGCTCGAGCGGGGCTCAACGCCCTCG |-----|
58 63
68 CAGACGCCGTAAAGTGTGACCTTGGCCCGAAGGCTGCAACGCTGTGAGAGAAGT |-----|
70 75 84 89 106 111
128 GGGGCGCCCCCAGATCACCAGATGCTGTGCTTCATGCGCAAGAGATCGAGCTGAGG |-----|
130 135 148 153 163 168
|-----|

```

188 ACCCGTAGAGAAATCGCGCTGAGCTGTGTCAAAGAGTTCGCCAAGAAGACCGAGANG
      205 210 226 231 241 246
      -|
      |-----|
248 TCGCGGGGAGACGGCACCAACACCGCCACCTGCTCCTCAGGCTCTGGTTCCGGAAGGCC
      253 258
      -|
      |-----|
308 TCGCAAGCTGCAGACCGCGCCCAACCCGCTCGGCTCAAGCGTGCATTCGAGAAGGCTG
      313 318 325 330
      |-----|
368 TCGAGGCTGTACCCAGTCTGCTGTGAAGTCGGCCAGAGAGTGCAGACCAAGAGACGA
      384 389
      |-----|
428 TTTCGCAACCGCGCGCATTTCCGCGGCGACACCCAGATCGCGAGCTCATTCGCCGAG
      441 446 454 459 478 483
      |-----|
488 CCATGACAAAGTGTGCGCAACGAGGGTGTATCACCGTTCGAGAGATC
      ...
563 CCGAGGTAATGCGCTTCGACAAAGGGTATCATCTCGGGTTACTGTGACCCAGCCGAGC
      613 618
      |-----|
623 GCCAGGAAGCCGTCCTTGAGAGATCCCTATCATCTGTGTAGCTCCAGAGTG
      625
      -|
692 TGGTCCCGGCTGTGAGAGATGATCCAGGCGCGCAAGCCGCTGCTATCATTCGCCGAGG
      742 747 751
      |-----|
752 ACCTCGAGGCGAGGCGCCCTGTCCACGCTGGTGTCAACAAAGATCCGGGCACTTCAAGT
      756
      -|
812 CCGTCCGCTCAAGGTCGCGGCTTCGGTGACCGCCCAAGCGCATCTCAGAGACATGG
      814 819 852 857
      |-----|
872 CCATCTCAACCGGTGTCAGAGTGTCAAGCAAGATCGGGCTGTCCCTGGAACCGCCG
      892 897 931
      |-----|
932 ACCTTCGCTGTGCGGCCAGGCGCCGCAAGTCTCTGTCAACCAAGAGACCAACCATTCG
      936 961 966 988
      -|
992 TCGAGGGCTTCGGGCGATTCGGATCGCCGCGGGGGTGTGAGATCCGCGCCGAGA
      993 1003 1008 1018 1023
      |-----|
1052 TCGAAGACAGCGACTCCGACTACGACCGGAGAGCTGAGAGCTGAGCCCAAGCTGG
      1060 1065 1094 1099
      |-----|
1112 CCGGCGGCTTTCGGGTGATCAAGCGCGGAGCTGCCACCGAGTGGAGTCAAGAGCCGA
      ...
1172 AGCACCAGCATGAGAGCGCGCTCCGCGCAACGCGAAGGCTGCCGTCAAGAGGGGAGTGTG
      1186 1191 1225 1230
      |-----|

```

```

1232 CCGGTGGCGGCTGGGCTGTGTCGAATGCGCTCTGCGGCTGACACCGCTGACCG
      1228
      -|
      |-----|
1232 CCGGTGGCGGCTGGGCTGTGTCGAATGCGCTCTGCGGCTGACACCGCTGACCG
      1273 1278 1291
      -|
      |-----|
1292 GCGAGAGCCACCGGTGCCAATCGCGCTGCGCTGTGCGGCTGCAAGACGA
      1296 1315 1320 1326 1331 1351
      |-----|
1352 TCGCTTCACAGCGCGGCTGTGAGCCCGGCGCTGTGCGGCTGCAAGGCTGTCCACTGCCCC
      1356 1378 1383 1381 1386
      |-----|
1412 CCGGTACAGGCTTCACACCGCGGAGCGGTGAGTACGAGACCTGTCTAAGGCGGCGTGC
      1426 1431 1465 1470 1468
      |-----|
1472 CCGACCCGCTGAGAGTACACCCGCTCGGCTGCAGAACGCGGCTCATTCGGGCTGTGT
      1473 1497 1502 1512 1517
      -|
      |-----|
1532 TCTTCACACCGAGGCGCGCTGTCGCGCCGACAGCCGAGAGAGCGGCTCCGCGGCGG
      1549 1554 1572 1577 1588
      |-----|
1592 -|
      1593
      |-----|

```

32 matches found in sequence:
aaz11372: Nucleotide sequence of *M. vaccae* antigen GV-27B.
(from "mycobacterieng.seq")
TOIG of: aaz11372 check: 1007 from: 1 to: 985

ID AAZ11372 standard; DNA: 985 BP.
AC
XX AAZ11372:
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of *M. vaccae* antigen GV-27B.
XX
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN MO9932634-AZ.
XX
PD 01-JUL-1999.
XX
XX 23-DEC-1998; 98MO-NZ00189.
XX PF
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.


```
PR 23-DEC-1997; 97US-0997362..
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
XX DR P-PSDB; AAY14911.
XX
XX Enhancing immune response to an antigen
XX
XX Claim 3; Page 216; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 403 BP; 60 A; 142 C; 134 G; 67 T; 0 other;
XX
AAZ1373 Length: 403 March 5, 2002 14:19 Type: N Check: 2096
Found using 'seq2-3' (pappu403.key)
1 GGATCCGGCGACCGCGCTGTGACCACTACACCGCGCGCACTGGAGCGCGCA
22 27
61 GAATCCGCTCGCATCG
...
106 GGAAGGCCCGCGGCGTATAGTTCTGTGCGCGCACCAACAGGTGTGACCTGTTACCG
156 161
166 CCACACGGCGCGCTTCGACACCGGAGAGTCCCTGGCTGCGGTTTTCGACCGAATTTC
172 177
226 GT
...
300 TCACGACGGGCGTGTCTCCCTCGCCGAGCAGCAAGGTTCTGGCGGACGCGGACG
350 355
356
-1
-1
360 ACATCGCGCGGCTTCCTCCGCGACGCGGCGCGCTTCTCCGCC
362 367
361
-----
7 matches found in sequence:
aaz1374; 3' nucleotide sequence of M. vaccae antigen GV-29.
(from "mycobacterieng.seq")
```

```
TOIG of: aaz1374 check: 9908 from: 1 to: 336
ID AAZ1374 standard; DNA: 336 BP.
XX
XX AC AAZ1374;
XX
XX 25-OCT-1999 (first entry)
XX
XX 3' nucleotide sequence of M. vaccae antigen GV-29.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX W09932634-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-N200189.
XX
XX 04-DEC-1998; 98US-0205426.
XX
XX 23-DEC-1997; 97US-0996624.
XX
XX 23-DEC-1997; 97US-0997080.
XX
XX 23-DEC-1997; 97US-0997362.
XX
XX 11-JUN-1998; 98US-0095855.
XX
XX 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
XX DR P-PSDB; AAY14912.
XX
XX Enhancing immune response to an antigen
XX
XX Claim 3; Page 216; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 336 BP; 45 A; 110 C; 132 G; 49 T; 0 other;
XX
AAZ1374 Length: 336 March 5, 2002 14:19 Type: N Check: 9908
Found using 'seq2-3' (pappu403.key)
...
14 CGGCGCGCGGAGTTGACCGCGCGAGAAAGCGGTGAGAGCCACGACGCGGATA
64 69
74 CCGGGACCTGTGCTCTACGACGCGTGCAGCGGGGTGACGCTCTCCGTGCGGTGACGCT
-----
```



```
111 116
134 GTGAGAGTGCAGCTGCGCCGCTGGTGCACCGGGTGTGGCCGGACAGCCGTTGCTGGCGCT
144 149
147 152
194 GGAGCGCATGAGATGGAGACCGTGTCTGCGCCCGCCCGCGACGGGGTGTACACCCAGA
197 202
...
276 GTGCATCCCGGACCCCACTGTGTGTGTGGACCGAGTCCGCGATGAGCCCGCTGC
326 331
336 A
-----
14 matches found in sequence:
aaz11377 ; Nucleotide sequence of M. vaccae antigen GV-24B.
(from "mycobacterng.seq")
TOIG of: aaz11377 check: 9748 from: 1 to: 1111
ID AA11377 standard; DNA: 1111 BP.
XX
XX AA11377;
XX
XX 25-OCT-1999 (first entry)
XX
XX Nucleotide sequence of M. vaccae antigen GV-24B.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
XX
XX WO9932634-A2.
XX
XX 01-JUL-1999.
XX
XX 23-DEC-1998; 98WO-NZ00189.
XX
XX 04-DEC-1998; 98US-0205426.
XX 23-DEC-1997; 97US-0996624.
XX 23-DEC-1997; 97US-0997080.
XX 23-DEC-1997; 97US-0997362.
XX 11-JUN-1998; 98US-0095855.
XX 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
XX WPI: 1999-430163/36.
XX
XX P-PSDB: AA114913.
XX
XX Enhancing immune response to an antigen
XX
XX Claim 3; Page 218; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
```

```
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
XX Sequence 1111 BP; 235 A; 369 C; 339 G; 168 T; 0 other;
XX
XX AA11377 length: 1111 March 5, 2002 14:19 Type: N Check: 9748
XX Found using 'seq2-3' (pappu403.key)
...
3 CCGACAGTGGGACCTCGACACACAGTCACAGAGAGCGGCCCGCCAGCGGCCCTGCG
53 58
63 GCGTCTCCAACTGGCCGCTCTATATGCGCGACGGTTTCATCGACCGTTCCAGACCGGCT
106 111
123 CGGCGATCAGCGGTGACTACTACAAGAAGACTTCACAGACGACGAGTGTTCGCCAAGG
155 160
183 TCAAGAGACCTTGTGCGGCAAGCAGACATAGCGCGGACCTGTGTATCCCAACGAGT
215 220
243 TCATGCGCGCGCGGTCAAGGCGCTGGG
...
390 TCGGTCTGCGCTACACAAAGAGCGACCGACCGGATATCCGACCATTCAGACGACTCT
440 445
450 GGGATCCCGCGCTTCAAGGCGCGCGTCACTCTTCTCCAGCTCAGAGCGGCTCGCA
488 493
510 TGATCATGCTCTCCAGGGAAGTCTGCGCGAAGATCCGACCGAGTCCATTCAGCAGG
593 598
570 CGGTTCATCTGTCCGCGAAGACGACGAGGGGTCAATCCGTTCACCGGCAC
613 618
630 GACTACGCGCGACGACTGCGCGCGGCAAGAAACATCGCATCGCGGCGGTACTCGGTGACG
632 639 644 659 664 686
690 TCGTTCAGCTTCAGAGCGGACCAACCCCGATCTGCAATTCATTCGCGGATTCGCGGCGG
728 733 746
750 ACTGTCTGTGACACAGATGTGTATCCCGTACACACGACGAGACCAAGAGC
751
...
32 matches found in sequence:
aaz11378 ; Extended DNA sequence for M. vaccae antigen GV-38A.
(from "mycobacterng.seq")
TOIG of: aaz11378 check: 1477 from: 1 to: 1420
ID AA11378 standard; DNA: 1420 BP.
```

[illegible]

| | | | |
|-------------------------------|--|------|------|
| 1266 | CGCAGCGGCTGAGAGAACTACCGTGCCTGACACGATGCGCCCGTACAGACATCGACCGCTTCG | 1317 | 1372 |
| 1326 | TGCACCGAAAGCCGATCCTGCTGCACGATCGGGGCCGCGATGTCGCCACCGCGCGCGCC | 1367 | 1372 |
| 1386 | ACGACTTCGGTGTGATGGCGGACTCGCAGGACTGA | | |
| ----- | | | |
| 41 matches found in sequence: | | | |
| ID | AA211379 standard; DNA; 2172 BP. | | |
| XX | | | |
| AC | AA211379; | | |
| XX | | | |
| DT | 25-OCR-1999 (first entry) | | |
| XX | | | |
| DE | Nucleotide sequence of M. vaccae antigen GV-38B. | | |
| KW | Mycobacterium vaccae protein; antigen; T cell activation; cytokine; | | |
| KW | dendritic cell maturation; infectious disease; immune disorder; cancer; | | |
| KW | respiratory system; mycobacterial infection; allergy; tuberculosis; | | |
| KW | leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis; | | |
| KW | dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma; | | |
| KW | squamous cell carcinoma; melanoma; ss. | | |
| XX | | | |
| OS | Mycobacterium vaccae. | | |
| XX | | | |
| PN | W09932634-A2. | | |
| XX | | | |
| PD | 01-JUL-1999. | | |
| XX | | | |
| PF | 23-DEC-1998; 98WO-NZ00189. | | |
| XX | | | |
| PR | 04-DEC-1998; 98US-0205426. | | |
| PR | 23-DEC-1997; 97US-0996624. | | |
| PR | 23-DEC-1997; 97US-0997080. | | |
| PR | 23-DEC-1997; 97US-0997362. | | |
| PR | 11-JUN-1998; 98US-0095855. | | |
| PR | 17-SEP-1998; 98US-0156181. | | |
| XX | | | |
| PA | (GENE-) GENESIS RES & DEV CORP LTD. | | |
| PI | | | |
| XX | Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J; | | |
| DR | WPI: 1999-430163/36. | | |
| XX | P-PSDB: AAY14915. | | |
| XX | | | |
| PS | Enhancing immune response to an antigen | | |
| XX | Claim 3; Page 221-222; 243pp; English. | | |
| XX | | | |
| CC | The invention provides heat-killed Mycobacterium vaccae, or recombinant | | |
| CC | M. vaccae proteins. The M. vaccae proteins may be employed to activate | | |
| CC | T cells and natural killer cells, to stimulate the production of | | |
| CC | cytokines, to enhance the expression of co-stimulatory molecules on | | |
| CC | dendritic cells and monocytes, and to enhance dendritic cell maturation | | |
| CC | and function. The proteins can be expressed by standard recombinant | | |
| CC | methodology. Pharmaceutical compositions comprising the proteins or | | |
| CC | nucleic acid sequences encoding the proteins can be used for the | | |
| CC | treatment, prevention, and detection of disorders including infectious | | |
| CC | diseases, immune disorders and cancer. In particular, the compounds and | | |
| CC | methods are used for treatment of diseases of the respiratory system, | | |
| CC | such as mycobacterial infections, asthma, allergies, tuberculosis, | | |
| CC | leprosy, sarcoidosis and lung cancers, and disorders of the skin such as | | |
| CC | psoriasis, atopic dermatitis, eczema, allergic contact dermatitis, | | |
| CC | alopecia areata, and skin cancers such as basal carcinoma, squamous cell | | |
| CC | carcinoma and melanoma. | | |

```

XX      SQ      Sequence 2172 BP: 387 A; 694 C; 721 G; 370 T; 0 other:
AAZ11379 Length: 2172 March 5, 2002 14:19 Type: N Check: 4735
Found using 'seq2-3' (pappu403.key)

1      TAGATGACAAATTCTGCGCTGGAAATGGCGGACGACCTCTGAAACACCCGACCGAAMAAAGACGC
      |-----|
      30      35

61      |-----|
      GGGCGGTACCAACCTCTCTGCGGATGAGCATCCAGTCCAAAGTGGTGTGATGCTGTG
      62      67

...

259      GACCTGAAACAACTCGATGATGATTACTCGCGGGGCGGACGACATGCCACGAGAGGCATCGGC
      |-----|
      309      314

319      |-----|
      GCGTTACGCGAGCGTTTCCGTCACGCTCGGGGATGCGAGCATCAATACCGGCGGACGCGCG
      325      330      346      351      359      375

379      |-----|
      TCATTGCGCGGTTACTACAGACCGGACGTTGCGCAACAGCAGCACCCTTGACGACGCGGAAC
      380      402      407      424      429

439      |-----|
      CCGGTGAGAGTCCGCGCGCTATCCGCAAAATCCAAACCCCGACGCGTATCTGACGCGCTC
      445      450      479      484      492      497

499      TATACCCCGCGCTTTCAGAACTGGGAGAGAAAGCGATCGCGTTCGAGAGCGGCGGACGCGC
      |-----|
      528      533      541      546

559      |-----|
      AGGCGCTGCGTGGCGCGCCAAATGCGAATTCACAGATCTTCCGGAGATCGTGCA
      559      564

...

713      GCCCTTATCGCAACCGGGAACGTTCGGAACCTTACGAGAAGGCGGT

653      AGGGCAACGTGGTGTACTCCGCTTACAAAGGGGCGGAGTCTCGGAGCAAAACATCGTCAAG
      |-----|
      703      708

...

995      TGGCGTTCGACTCCCGGCTGTTCCGCGAGAAACGGGAGAAAGTTCTTGCGCGGACGTCGTCG
      |-----|
      1045      1050      1048      1053

1055      AGGGGGGAACCCCGCGGAGGTTCGCGCAAGATCGGTTGACCGCGCGGCAACGACGCTGG
      |-----|
      1075      1080

1115      TGCAGCGCGGTGACCAACCGCTCGCTCGAGAGAGGCCCAACGCGGCAACACCGGGAGGACGA
      |-----|
      1167      1172      1167      1170

1175      TCGAGGAGCAGCTATCTCGCGGACGAGGCGTTACAGGCGGTATCTACCGGTTGACCTGCGCG
      |-----|
      1180      1185      1200      1205

1175      |-----|

```

```

1235 GACTGCACCTGGTGTATCTGGCCAGATGCACACCGAGGCGTTGCGCCCGGTGGGCGC
1275 1280
1295 AGTTACCAAGACCCCTGTGCTGTCTGCAGCGTGATCATCATCTTCGGCGTGTGGCGG
1344 1349
1355 CCATGCTCTGCGCGGCTTGTTCGTCCGCTCCGATCCGGGTTGCAGGCCCGCCGACG
1405 1410
1415 AGATCAGCGCGGTGACTACCGCTCGCTGCGGTTGTCTGTCGTGACGAATTCGGCG
1471
1475 ATCTGACACACAGCTTTTCAAGACATGAGTCCGAATCTGTGATCAAGAGCAGCTGTCTG
1476 1492 1497
1535 GCGAGAGCGCGCCGAGAACCAAGGCTGATGTCTGCTCGATGCGCCGACCGGTGATGC
1595 AGCGCTACCTTCGACGCGGAGGAGACGATCGCCGAGACACAGAGAGCGGTGATCT
1595 1600 1617 1622 1621 1626 1639 1644
1655 TCGCGCATGATGAGGCTCGACGAGTGTTCGCGCATGTTGACCTCCGAGAACTGATGG
1715 TGTGTGTCACGACCTTGACCCGCGCAGTTTCGACGCGCGCCGAGAGTCTCGGGTGCAGC
1723 1728 1744 1749
1775 ACGTGGCGAGCTTCACAGAGGAGGTTGACCTGCGCACTCGGGTTAGCGGTGCCGCGCTGG
1782 1787
1835 ACACGTCGCGCGCAGGTCATATTTGCGATCGAATGAGACCGCATCATGACCGGCA
1837 1842
...
1961 TGGTGGGGCGTCCACCTTGGCGTACGACATGTGGGTTGCGCGGTGATGCTGATACC
2011 2016
2021 AGGTGACGCGCGCTCCCGCCAGCCGCGCATCTACGTCACCTCGCGGGTCCAGAGGTCA
2081 TGCAGMAACTCTGCACTTCTGCGCGCGCGGAGGTGTCGCGGAGACGCGGCTGAGA
2101 2106 2116 2121 2131 2136
2141 CGGTCTGGCGGTTGCAGGCGCCACCGGATGA
2165 2170
-----
23 matches found in sequence:
aaz11380 ; Extended DNA sequence for truncated M. vaccae antigen GV-38A.
(from "mycobacterieng.seq")
TOIG of: aaz11380 check: 9477 from: 1 to: 898
ID AA211380 standard; DNA; 898 BP.
XX
AC AA211380;
XX
DT 25-OCT-1999 (first entry)
XX
DE Extended DNA sequence for truncated M. vaccae antigen GV-38A.
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;

```

```

KM respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
OS Mycobacterium vaccae.
PN W09932634-A2.
PD 01-JUL-1999.
PF 23-DEC-1998; 98WO-N200189.
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
PA (GENE-) GENESIS RES & DEV CORP LTD.
PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX MPI: 1999-430163/36.
XX P-PSDB; AAY14916.
XX
XX Enhancing immune response to an antigen
XX
XX Claim 3; Page 224; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 898 BP; 159 A; 298 C; 307 G; 134 T; 0 other:
XX
XX AA211380 Length: 898 March 5, 2002 14:19 Type: N Check: 9477 ..
XX Found using 'seq2-3' (pappu403.key)
1 GAGCAACCGTTCCGGCTCGCGCATGATCACGCTCCCAACCGCGCGCGCGCTCC
19 24
61 GCCCAGCGCCGCT
...
86 ACTGGCGTGACACATATGACACCGCGCAACTGTGTATATGCCCAACGCGCAAC
136 141
146 TCGCGCGCGCTGTGTTACCAATTTACAGCGCGCGTGGAGAGACCGGCTGACCGTGC
156 161
206 TCAACCACTTCAACGCGCGGAGACCCCGCATATGTCTCGAGATGCTGTCTGCTGCG
207 217 222 255 260

```

[illegible]

```

PD 01-JUL-1999.
XX
XX PF 23-DEC-1998; 98WO-NZ00189.
XX
XX 04-DEC-1998; 98US-0205426.
XX 23-DEC-1997; 97US-0996624.
XX 23-DEC-1997; 97US-0997080.
XX 23-DEC-1997; 97US-0997362.
XX 11-JUN-1998; 98US-0095855.
XX 17-SEP-1998; 98US-0156181.
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
DR P-PSDB; AAY14917.
XX
XX Enhancing immune response to an antigen
XX
PS Claim 3; Page 224-225; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
XX Sequence 2013 BP; 355 A; 646 C; 678 G; 334 T; 0 other:
SQ
AAZ11381 Length: 2013 March 5, 2002 14:19 Type: N Check: 3884 ..
Found using 'seq2-3' (pappu403.key)
...
100 GACCTGAAGAACTCGATGCTGATTACTCGCGGGCAGACACTGCCACGGAGCGGATCGGC
150 155
160 GCGTTACAGCAGCGGTTTCCGTACAGTCGGCGCATCGCAGCATATACCGGGCAGCGCGG
166 171 187 192 200 216
220 TTAATTGCGCGCTTACTACGACCGGACGTTGCCACACACACCTCGACGACGGAAAC
221 243 248 265 270
280 CCGCTGACAGTCGCGCGCGCTCATCCGAAATCCAAACCCGACGCGTATTCGACGGCGTC
286 291 320 325 333 338
340 TTAATCCCGCGCTTTCAGAACTGGGAGAAGCGGATCGCGTTTCAGACGACGCGCGACGCG
369 374 382 387
400 AACGGCTGTGGTGGCGCGCAATGCAGATTCACAGCATGTTCTTCGCGGAGATCGTCA
400 405
...

```

```

494      AGGGCAACGTGGTACTCCGCTACAAAGGGCCGGAATCTCCGGACAAACATCTCAACG      544 549
554      GCCCCATTCGAACCGGGAACGTGTGGAAAGCCTACGAGAGGGCGT
...
836      TGGCTTCGAGACTCCCGGCTGTTCGCGGAGAAACCGGGAGAAAGTTCTTGGCCGACGTCGTCG      886 891
889 894
896      AGGGGGGAAACCCCGCGAGGCTCGCCGACGAATCGTTGACCGCGCGGCAACACGCTGG      916 921
956      TGCAGCCGGTGCACCAACCGGCTCCGTCGAGAGACCCCAACGCGGCAACACCGGCAACGAGCA      1008 1013
1016      TCGAGGACGACTATCTCGGCGCCACGAGCGGTACAGGCGTACTCAACCGGTGACCTGCGCGG      1021 1026
1041 1046
1076      GACTGCACGTGGTGTATGTCGGCCAAAGATGCACACCGAGCGGTTCGCGCCCGGCGGCGGC      1116 1121
1136      AGTTACACGAGACCCCTGCTGCTGTGACGAGGTATCATCTTGGCGGTGCTGCGCGG      1185 1190
1196      CCATGCTGTGGCGCGGTTGTGTCGTCGTCGATCCGCGGTTGCAGGCGCGGCCCGCCAGC      1246 1251
1256      AGATCAAGCGCGGTGACTACCGCTCGCTCGCTGCGGTGTGTCGTCGTCGAGCAATTCGCGG      1312
1316      ATCTGCACACAGCTTTTCACAGCATGAGTCGCAATCTGTGATCAAGAGACAGCTGCTCG      1333 1338
1376      GCGAGGAGCGCGCGGAGAACCAACGCGCTGATGCTGTCCCTGATGCCCGAACCAGTATGC
1436      AGGCGTACCTCGACGGGAGAGAGATCGCCCGAGGACCAACAGACGTCACGGTATCT      1458 1463
1462 1467
1496      TCGCCGACATGATGGGCTTCGACGAGTGTCCGCGCATGTTGACCTCCGAGGAACGTATGG
1556      TGGTGTCAACGACGCTGACCGCCGCGAGTTCGACGCGCGCGCGAGAGTCTCGGGTGCAGAC      1564 1569
1585 1590
1616      ACGTGGGAGACGCTGCACGACGAGGTACCTGGCCAGCTGCGGGTTAGCGCTGCCCGGCTG      1623 1628
1676      ACACGTCGCGGCGCAACGCTCAATTCGCGGATGCAATGAGACCGCATCATCGACCGCA      1678 1683

```

```

1802      TGGTGGGGCGGTCACGTTGGCTACGACATGTGGGGTTTGGCGGTGCTGATGTCGTAAC      1852 1857
1862      AGGTGACAGCGCGGCTCCCGCCAGCCCGCATCTACGTCACATCTGGCGGTGCACGAGGTCA
1922      TGCAGGAAACTCTGACACTTCGTCGCCCGCGGAGAGTCTGCGCGAGCGCGGCTGAGA      1942 1947
1957 1962
1982      CGGTCTGGCGGTTGCAGGGCCACCGCGCATGA      2006 2011
-----
10 matches found in sequence:
aa211382 ; Nucleotide sequence of M. vaccae antigen GV-41.
(from "mycobacterieng.seq")
TOIG of: aa211382 check: 8110 from: 1 to: 520

ID      AA211382 standard; DNA; 520 BP.
XX
XX      AC
XX      AA211382;
XX
XX      25-OCT-1999 (first entry)
XX
XX      DT
XX      DE
XX      Nucleotide sequence of M. vaccae antigen GV-41.
XX
XX      Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX      dendritic cell maturation; infectious disease; immune disorder; cancer;
XX      respiratory system; mycobacterial infection; allergy; tuberculosis;
XX      leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX      dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX      squamous cell carcinoma; melanoma; ss.
XX
XX      OS
XX      Mycobacterium vaccae.
XX
XX      W099932634-A2.
XX
XX      PD
XX      01-JUL-1999.
XX
XX      PF
XX      23-DEC-1998; 98MO-NZ00189.
XX
XX      PR
XX      04-DEC-1998; 98US-0205426.
XX      23-DEC-1997; 97US-0996624.
XX      23-DEC-1997; 97US-0997080.
XX      23-DEC-1997; 97US-0997362.
XX      11-JUN-1998; 98US-0095855.
XX      17-SEP-1998; 98US-0156181.
XX
XX      PA
XX      (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX      PI
XX      Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX      WPI; 1999-430163/36.
XX      DR
XX      P-PDB; AAY14918.
XX
XX      PT
XX      Enhancing immune response to an antigen
XX
XX      PS
XX      Claim 3; Page 228; 243pp; English.
XX
XX      The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX      M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX      T cells and natural killer cells, to stimulate the production of
XX      cytokines, to enhance the expression of co-stimulatory molecules on
XX      dendritic cells and monocytes, and to enhance dendritic cell maturation
XX      and function. The proteins can be expressed by standard recombinant
XX      methodology. Pharmaceutical compositions comprising the proteins or
XX      nucleic acid sequences encoding the proteins can be used for the
XX      treatment, prevention, and detection of disorders including infectious
XX      diseases, immune disorders and cancer. In particular, the compounds and
XX      methods are used for treatment of diseases of the respiratory system,
XX      such as mycobacterial infections, asthma, allergies, tuberculosis,

```



```

421 AACATCGTCT
...
482 CGTTCCGGGTCTGGCCGACCGGGTGGCGGACAGCCCTTACTGTCGTAACGCCAAG
532 537
542 TGGTGTCTACAACTCGTGGCCAGTTTCAGACCCAGAGACGATGACGCCGCTTCG
582 587
602 TCGACAGCCAGAGCTGCGGCGTGGCGTCCACCGAGCCGCTGCGCCACTTGGCG
626 631 642 647
662 GAATGCCGTCTGCTGATCGAGGGCACCTACCGGAGACACATGAAGCTGAACAGT
722 CCGGCGCAGCTCATTTGCCACCGGGGGCCGACCACTACTGCTGCTGCTGCA
725 730 768 773
782 CCACACGCTCGACAGGCGCTGGCCGAAAGCCGGAGAGCCACCGAGCGATTTGCAAG
787 792
842 GCTTCAAGGTGAGGCTTCCGGGTCGGGTCGGCCGACCGCACCTGACCCGGTGCC
853 858
902 CCGGTGTCGGCCCGCCCGGGCGGCGTGGCCGCTGGCCGACACCCCGG
922 927 935
962 CTCGCCGTGTCGGCGCGTGGCGCCGACACAGCTGCTGGGACTGAGGAGTACGCT
981 986 1017 1020
1022 CGTGTGTCGGCGGCGGAAAGCTGGCGCGGGGAGCAGCGCCCTTCT
1022 1044 1049 1055 1060
1023 1028
-----
3 matches found in sequence:
aaz11384 : M. vaccae antigen GV-40 partial gene sequence.
(from "mycobacterin.seq")
TOIG of: aaz11384 check: 6825 from: 1 to: 207
ID AAZ11384 standard; DNA: 207 BP.
XX
AC AAZ11384;
XX
DT 25-OCT-1999 (first entry)
DE
XX M. vaccae antigen GV-40 partial gene sequence.
XX
XX Mycobacterium vaccae protei; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma; SS.
XX
OS Mycobacterium vaccae.
XX
XX WO9332634-A2.
PN

```

```

XX 01-JUL-1999.
PD
XX
XX 23-DEC-1998; 98WO-N200189.
PF
XX
XX 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
XX P-PSDB: AAV14920.
XX
XX Enhancing immune response to an antigen
XX
XX Claim 3: Page 230; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of disorders of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
XX Sequence 207 BP; 40 A; 66 C; 66 G; 35 T; 0 other;
XX
XX AAZ11384 Length: 207 March 5, 2002 14:19 Type: N Check: 6825 ..
XX Found using 'seq2-3' (pappu403.key)
XX
XX
XX 29 CCGGCCGATCCCGCGAGTACATCCGTCGTGATGCCGGCGCAGAGCCCATGC
79 84
XX
XX 89 AGTAGCGGTGCTGGCGCGTACCCGCTGTTACGTCACGTCGCTGTCAGCGTGS
121 126 132 137
XX
XX 149 CTAACACGAGTCTGACTGTCGGAATGGCATTTCAAG
XX
XX
XX 20 matches found in sequence:
XX aaz11385 : Nucleotide sequence of M. vaccae antigen GV-44.
XX (from "mycobacterin.seq")
XX TOIG of: aaz11385 check: 7213 from: 1 to: 898
XX
XX ID AAZ11385 standard; DNA: 898 BP.
XX
XX AC AAZ11385;
XX
XX DT 25-OCT-1999 (first entry)
XX
XX Nucleotide sequence of M. vaccae antigen GV-44.
XX
XX

```


KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN W0932634-A2.
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98MO-N200189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI: 1999-430163/36.
XX
DR P-PSDB; AAY14921.
XX
PT Enhancing immune response to an antigen
XX
PS Claim 3; Page 231; 243pp; English.
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancer, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 898 BP; 189 A; 316 C; 274 G; 117 T; 2 other;
AAZ11385 Length: 898 March 5, 2002 14:19 Type: N Check: 7213
Found using 'seq2-3' (pappu403.key)
...
46 TAGTCAGCGCACACTTCAACCGAATAATGAGAGACAGAGACACCCGCTACGATCCGT
96 101
106 GTTGTGTGAACGGCTTTCGGCCGTATCGAGCACTTCTTCGGCGCGCTGAGCGGCAG
166 AAGCGCAAGGCAAGAACAAGACATGAGATGTCGGGCTCAACGACTACCGACAC
196 201 208 213
226 --! GCCACGCTGGCGCACTGCTGAAGTTGACTCGATCTCGGCGGCTGCCCTACGACGTG
228

286 AGCCGTGAAGGCGAGACACCATCTGCTGGCAGCAGCAGCAAGATCAAGCGCTCGAGTTC
307 312 333 338
310 315
346 AAGGAAGCCCGCGCGCTGCTGGGCGGACCTGGCGCTGCGATCTGATGATTC
360 365 373 378 382 387 391 396
388 393 394 399
406 ACCGCACTTTCACCAAGCGCAGACAGGCCAGCGCACCTCGACGCGGCGCCAGAAAG
454 459
466 GTTCATCATCTCCGCGCGCGCCAGCAGATGAGAGATCATCAGCATCTGCTGGCGCTACAGAC
514 519 523
520 525
--! GACAAGTACGACGCGCAGCCAGAACATCATCTCCACGCGCTGTCACACCAAGATCGCTC
528
586 GCGCCGCTGGCGAAGGTATCATCAGCAGAGATTCGATCTCAAGGCGCTGATGACCAC
607 612 622 627
646 ATCCAGCGCTACACCCNGTCTCAGAACTTCA
...
836 TCACGACCTGACCGCGGAGCTGGGCAAGTCCGCAACCGTGGAGATCAACGCGCGCA
886 891
896 TGA

1 match found in sequence:
aaz11388; M. vaccae DNA fragment obtained by PCR amplification.
(from "Mycobacterieng.seq")
T0IG of: aaz11388 check: 4257 from: 1 to: 84
ID AAZ11388 standard; DNA: 84 BP.
XX
AC AAZ11388;
XX
DT 25-OCT-1999 (first entry)
XX
DE M. vaccae DNA fragment obtained by PCR amplification.
XX
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN W0932634-A2.
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98MO-N200189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.

```

PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
DR
XX
XX Enhancing immune response to an antigen
PT
XX
XX Example 17; Page 233; 243pp; English.
PS
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 84 BP; 21 A; 21 C; 25 G; 16 T; 1 other;
AAZ11388 length: 84 March 5, 2002 14:19 Type: N Check: 4257 ..
Found using 'seq2-3' (pappu403.key)

1 GNACTCATTTGACGTACTGAGAGAGCTGGGCTCGATGTCGCAAGCAGCGTCGGCA
|-----|
48 53
61 ATGAGAGACGTGTCACACCATTA

-----
5 matches found in sequence:
aaz11389; Nucleotide sequence of M. vaccae antigen GV-45.
(from "Mycobacterieng.seq")
TOIG of: aaz11389 check: 1491 from: 1 to: 337

ID AAZ11389 standard; DNA: 337 BP.
XX
AC AAZ11389;
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae antigen GV-45.
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
OS
XX W09932634-A2.
PN
XX 01-JUL-1999.
PD
XX
XX 23-DEC-1998; 98WO-NZ00189.
PF
XX
XX 04-DEC-1998; 98US-0205426.
PR
XX 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.

```

```

PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
DR
XX
XX P-PSDB: AAV14923.
DR
XX
XX Enhancing immune response to an antigen
PT
XX
XX Claim 3; Page 233; 243pp; English.
PS
XX
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 337 BP; 61 A; 107 C; 112 G; 56 T; 1 other;
AAZ11389 length: 337 March 5, 2002 14:19 Type: N Check: 1491 ..
Found using 'seq2-3' (pappu403.key)

1 GNACTCATTTGACGTACTGAGAGAGCTGGGCTCGATGTCGCAAGCAGCTCGGCG
|-----|
48 53
61 GTGAGAAATGTTTCGACACCATCGTGGCCCGCTGCACAGAGGTGAGAGGTACCATTC
|-----|
109 114
121 ACGGGCTTCGGTGTTCGACAGCAGCGTGGCCGAGCAGCGTGACCAATCCGGC
|-----|
143 148
146 151
181 ACCGGCAGACCGTGAAGGTCAAGCCCACTCAGTCCCGGATTCGTCGGCGGGGCTAG
|-----|
232 237
241 TTCAAGGCTGTTGTCTGTGGCGACAGAGACTTCGGCCGAGGGTCC
...
-----
20 matches found in sequence:
aaz11390; Nucleotide sequence of M. vaccae antigen GV-33.
(from "Mycobacterieng.seq")
TOIG of: aaz11390 check: 2495 from: 1 to: 1164

ID AAZ11390 standard; DNA: 1164 BP.
XX
AC AAZ11390;
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae antigen GV-33.
XX

```

KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KW squamous cell carcinoma; melanoma; ss.
 XX
 OS Mycobacterium vaccae.
 XX
 PN WO9932634-A2.
 XX
 PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-N200189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX WPI: 1999-430163/36.
 DR P-PSDB; AAI14924.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Claim 3; Page 234; 243pp; English.
 XX
 CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
 CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancer, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.
 CC
 SO Sequence 1164 BP; 165 A; 425 C; 397 G; 177 T; 0 other;
 AAZ11390 Length: 1164 March 5, 2002 14:19 Type: N Check: 2495 ..
 Found using 'seq2-3' (Pappu403.key)
 1 GGTGGCGCGATCGAGAGCGCGCCGCGGCTTCACGGCGGCTGATGATGTCGCGCGC
 18 23 38 43 60
 61 GCGCTGCGCTACGCGCTTCGGAGCGGCTCACTGTCGCGCGGGTCTGCTGC
 65
 233 GCATCAGATGATGCGGAGACCTCGAGCGGATGCTGTGCGGATCTGCGCAAGCGCCAT
 283 288
 293 CCCAGGGCAAGCGCGCGAGCCATCCGCTGGCGGAGCGCGGAA
 ..
 ..

437 GGAGCAACAGATGTTGCCCTTCACGGCGGCTCGACCGCAGCCATGACGACGATCCCGG
 487 492
 497 TGCCCGTGGAGGCTTCCCGCCGCTGGAGCGCGGTGTGATTCAGCACGACCACTACGACC
 507 512
 557 ACCTCGACATCGAACATGCTGCGCTTGGCGGACACCCAGCGGGCCCGTGTGGTGC
 574 579
 617 CGTTGGCATCGGCGCACACCTCGCAAGTGGGGCGTCCCGAGCGCGATCGTGCAGT
 649 654 667 672
 677 TGGAATGGCAGGAAGCCACCGCATATGACGACGCTGTGCTGCACCCCGCCGGC
 703 708 716
 737 ACTTTCGGAGCGTGTCTCTCCCGGACTCGACGCTGTGGCGCTCGTGGGTGTACCG
 768 773 777 782
 797 GCTCGTCGACAAAGCGCTTCTCGGTGGCGACCGGATACAGGAGCTTCGCGAGA
 810 815 823 828
 857 TCGCGCAGGATGACGGTCCGCTGATGACCTGTGTCGAGGCGGCTTACATCC
 859 864
 961 CATCTGACCTGACCGAGGTGACAAACAGCCTGATGTGCCATTCACATGGGCGCATTC
 1011 1016
 1021 CGCCTGCCCCGATCCGTTGCTCCGAGCCGCGGACGCTGCTGACCGCTGCCGAGCC
 1055 1060 1075 1080
 1081 GAGGGGTACGCTGACCGTGCAGATTCGCGTACGGGTGGACCCGAGTGGAGCTTC
 1134 1139
 1141 GACCCGTGTGGCGGTTTGAACC
 12 matches found in sequence:
 aaz11391: Nucleotide sequence of M. vaccae antigen GVC-13.
 (from "Mycobactering.seq")
 TOIG of: aaz11391 check: 2841 from: 1 to: 650
 ID AAZ11391 standard; DNA: 650 BP.
 XX
 AC AAZ11391:
 XX
 DT 25-OCT-1999 (first entry)
 XX
 DE Nucleotide sequence of M. vaccae antigen GVC-13.
 XX
 XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KW squamous cell carcinoma; melanoma; ss.
 XX
 XX Mycobacterium vaccae.
 OS
 PN WO9932634-A2.
 XX

PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-NZ00189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX WPI: 1999-430163/36.
 DR P-PSDB: AAV14925.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Claim 3; Page 235-236; 243pp; English.
 XX
 CC The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
 CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.
 CC
 SQ Sequence 650 BP; 98 A; 240 C; 215 G; 97 T; 0 other;
 AA211391 Length: 650 March 5, 2002 14:19 Type: N Check: 2841 ...
 Found using 'seq2-3' (pappu403.key)

531 CACTGGCCGTGATGTCGGCGGTGGTGTGATCTCCCGACAGATCTGTGTCGCGCGCGTC |-----|
 582 587
 591 ACATGTGTGGCGGCGCTGCGCTGGGTGGTGGTGGTGGCGGCGGCAAGATGAACAT |-----|
 602 607
 651

 13 matches found in sequence:
 aa211392; Nucleotide sequence of *M. vaccae* antigen GV-29.
 (from "Mycobacterium.seq")
 TOIG of: aa211392 check: 4645 from: 1 to: 743
 ID AA211392 standard; DNA; 743 BP.
 XX
 AC AA211392;
 XX
 DT 25-OCT-1999 (first entry)
 XX
 DE Nucleotide sequence of *M. vaccae* antigen GV-29.
 XX
 KW *Mycobacterium vaccae* protein; antigen; T cell activation; cytokine;
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KW squamous cell carcinoma; melanoma; ss.
 XX
 OS *Mycobacterium vaccae*.
 XX
 PN W09932634-A2.
 XX
 PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-NZ00189.
 XX
 PR 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
 XX WPI: 1999-430163/36.
 DR P-PSDB: AAV14927.
 XX
 PT Enhancing immune response to an antigen
 XX
 PS Claim 3; Page 237; 243pp; English.
 XX
 CC The invention provides heat-killed *Mycobacterium vaccae*, or recombinant
 CC *M. vaccae* proteins. The *M. vaccae* proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell

CC carcinoma and melanoma.
XX
SQ Sequence 743 BP; 107 A; 252 C; 268 G; 116 T; 0 other;
AAZ11392 Length: 743 March 5, 2002 14:19 Type: N Check: 4645
Found using 'seq2-3' (pappu403.key)

```
1  GGATCCGGGGACCCGGCTGGTGAACCAAGTACAAACCGCGCCGACCTGAGCGCCGA
22  27
61  GAATCCGTCGGCATCG
...
106  GGAGGGCCCGCGGCTATCAGTTCGTGGCCGACACCCAGAGTGTGAGTGTACCG
156  161
166  CCACAGGGCGGCTGGAAACCGGGAAGTCCCTGCTGTCGGGTTTTCAGCGAATTTC
172  177
226  GT
...
300  TCACCGACGAGCGGTCTCCCTCCGCGAGCAGCAGACGAGTCTGCGCAGACCGCAGC
350  355
-|-----|-----|-----|-----|
360  ACATCCCGCGGTTCCGTTCCCGGAGGCGCCGCTTTCGCCGAGCGAGCCGCGTG
362  367
361
...
420  CGGCGCGCGGCGAGTTCGACCGCGCGAGAAAGCCGTCGAAAGCCACCGCGCGATA
470  475
480  CCGGGGACCTGTGTCTCTACGACGAGTGCAGCGGGGTGCAGCGCTTCGCTGACGCG
518  523
540  TGTGGAAGTTCGACGTCGCGCGTGGTGACCGGGTGGTGCGGACAGCCGTTCTGCGCG
551  556
554  559
600  TGGAGCGAATGAATGAGAGACCGCTGCTGCGCCCGCCGCGACGGGGTGTACCCAGA
600  604  609
...
683  GTTCATCCCGGACACCCACTGTGTGTGTGCGGACCGGAGTGGCGCATGAGCGCGTGC
733  738
743  A
-----
10 matches found in sequence:
aaz11393 ; M. vaccae antigen GV-45 full length gene sequence.
(from "mycobacterieng.seq")
FOIG of: aaz11393 check: 2749 from: 1 to: 858
```

```
ID AAZ11393 standard; DNA; 858 BP.
XX
AC AAZ11393;
XX
DT 25-OCT-1999 (first entry)
XX
DE M. vaccae antigen GV-45 full length gene sequence.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN W09932634-A2.
XX
PD 01-JUL-1999.
XX
PE 23-DEC-1998; 98W0-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
XX WPI; 1999-430163/36.
XX DR P-PSDB; AAY14928.
XX
PT Enhancing immune response to an antigen
PS
PS Claim 3; Page 238-239; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 858 BP; 193 A; 269 C; 282 G; 114 T; 0 other;
XX
AAZ11393 Length: 858 March 5, 2002 14:19 Type: N Check: 2749
Found using 'seq2-3' (pappu403.key)
1  GAAATCCCGGCTGTAACCCCTCTTTCGCGGCCCTCAGACGGTAAGGGGCCAAG
31  36
61  CGGATTGAAAATATCTTCGCTGATGAGCCTGAAATTCGCGGCTGCTTGGAAATCAGA
120
121  GCGATGGGTTTACCGTGCCTAGTGTGTCGAAGAAGAGACACATGTTTCGGA
-----

```

```

...
125
183 CATGACAAGACAGCTCATCGACCTACTACTAGAGAAGCTGGCGCTCGATCGTCGGCA
|-----|
243 AGCGACTCGCGGGTGAGAACGTTGTCGACACCATCTGCGCGCGCTGCACAGGGTGA
243 248 262 267
303 GAGCGTCAACATCAAGGGCTTCGTGTTTCGAGCAGCGTCTGCGCGAGCAGCGGTGGC
|-----|
304 309 338 343 341 346
363 AGCGAATCCGGCGACCGCGAGACCGTGAAGTCAAGCCACCTCAGTCCGGCATTTCCG
|-----|
423 TCCCGGCGCTCAAGTTCAAAGGCTGTGTCTCTGCGGCACAGAAAGTTCCGGCGAGGTCC
427 432
...
760 AAGCGCGGACCAAGGCTGCACCGGCCAAGAAGGCTCGCGCCGCAAGAGCGCCGCC
|-----|
810 815
820 AAGGAAGCTCGCGGCAAGCGCGCGGAGCAAGTAAGTC
-----
10 matches found in sequence:
aaz11394: Nucleotide sequence of M. vaccae antigen GV-41B.
(from "mycobacterieng.seq")
TOIG of: aaz11394 check: 2416 from: 1 to: 570
ID AA211394 standard; DNA: 570 BP.
XX
AC AA211394:
XX
DT 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae antigen GV-41B.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; mycobacterial infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN W09932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998: 98WO-N20189.
XX
XX 04-DEC-1998: 98US-0205426.
PR 23-DEC-1997: 97US-0996624.
PR 23-DEC-1997: 97US-0997080.
PR 23-DEC-1997: 97US-0997362.
PR 11-JUN-1998: 98US-0095855.
PR 17-SEP-1998: 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI: 1999-430163/36.

```

```

DR P-PSDB: AAY14929.
XX
PT Enhancing immune response to an antigen
XX
PS
XX
XX Example 16; Page 239; 243pp; English.
CC The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 570 BP; 140 A; 186 C; 172 G; 72 T; 0 other;
AA211394 Length: 570 March 5, 2002 14:19 Type: N Check: 2416
Found using 'seq2-3' (pappu403.key)
...
23 CCTCTTCCATGCCGAGAGATGAGAAAGCCGTCTCGGTGGCACCAGACCTG
|-----|
73 78 81
83 CGTGAATGCTACCGCGCGCGGGAACCCCGCATCTTCAACCGATCAATCACTACT
|-----|
86
143 ACGCGGCTCCACCCCGATCAGCAGCTGTCCAGCATCAAGTGGCCGAGCGCGCAT
|-----|
145 150
...
221 AGCGAGCAGCTGCGCTTCATCGAGATGCGATCCGACTCGGCGCTGCAATC
|-----|
271 276
281 CGACCAAGAGCGGACATCATTCGGGTGTCGATCCGCAAGCTCACCAGAGAGCGCGCC
|-----|
286 291 332 337
341 GCGACCTGTCAAGCAGGCGCAAGGCCAAGGGCGAGGCGCAAGTGTGTCGCAACA
|-----|
376 381
401 TCCGTGCAAGCGGATGAGAACTCTCCGGATCAAGAGAGAGCGGAGCGCGGAG
|-----|
411 416 445 450 448 453
461 ACCAAGTGAACCGCGCGGAGAGAGATCTGCAAGACACCA
-----
37 matches found in sequence:
aaz11395: M. vaccae antigen GV-44 full length gene sequence.
(from "mycobacterieng.seq")
TOIG of: aaz11395 check: 1938 from: 1 to: 1364
ID AA211395 standard; DNA: 1364 BP.

```

XX AA211395;
AC 25-OCT-1999 (first entry)
DT
XX
XX M. vaccae antigen GV-44 full length gene sequence.
DE
XX
XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KM dendritic cell maturation; infectious disease; immune disorder; cancer;
KM respiratory system; mycobacterial infection; allergy; tuberculosis;
KM leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KM dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KM squamous cell carcinoma; melanoma; ss.
XX
OS Mycobacterium vaccae.
XX
PN W09932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;
PI
XX
XX WPI: 1999-430163/36.
DR P-PSDB; AAY14930.
XX
XX
XX Enhancing immune response to an antigen
PT
XX
XX
XX Example 16; Page 240-241; 243pp; English.
PS
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC T cells and natural killer cells, to stimulate the production of
CC cytokines, to enhance the expression of co-stimulatory molecules on
CC dendritic cells and monocytes, and to enhance dendritic cell maturation
CC and function. The proteins can be expressed by standard recombinant
CC methodology. Pharmaceutical compositions comprising the proteins or
CC nucleic acid sequences encoding the proteins can be used for the
CC treatment, prevention, and detection of disorders including infectious
CC diseases, immune disorders and cancer. In particular, the compounds and
CC methods are used for treatment of diseases of the respiratory system,
CC such as mycobacterial infections, asthma, allergies, tuberculosis,
CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC carcinoma and melanoma.
XX
SQ Sequence 1364 BP; 278 A; 470 C; 428 G; 188 T; 0 other;
AA211395 Length: 1364 March 5, 2002 14:19 Type: N Check: 1938 ..
Found using 'seq2-3' (pappu403.key)

106 GTTGTGTGAACGGCTTCGGCCGTATCGACGCAACTTCTCCGCGCGCTGGACGGCGAC
46 TAGTCGACGGCACACTTCACCGAAAAATGAGGACAGAGAGACACCCCTGACGATCCGT
166 AAGGCGGAAGGCAAGAAAGACATCGAGATCGTCGGCGTCAACGACCTCACCAGCAAC
196 201 208 213 223

226 --|
GCCACGCTGGCGACACTGCTGAAGTTGACTGCATCTCGGGCGGCTGCCCTACGACGTC
228
286 AGCCTCGAAGGGGAGAGACACCATCGTCGTGGGCGAGCACAAGATCAAGGCGCTGAGGTC
307 312
310 315
333 338
346 AAGGAAGCCCGCGCGCTGCGCTTGGGGGAGACTGGGCTCCACGTCGTCGAGTCC
360 365 373 378 382 387 391 396
388 393 394 399
406 ACCGCGATCTTCACCAAGCGCGACAAAGGCCAGGCGCACCCTCGACCGCGGCCCAAGAG
454 459
466 GTTCATCATCTCCGCGCGCGCGCCACCGAGATGAGAGACATCACCATCGTCTCGGCGTCAGAGAC
514 519 523
520 525
526 --|
GACAGTAGCAGAGCGCAGCGCAGAACATCATCTCCACGCGTGTGACACCAAGACTGCGTC
528
586 GGCCCGCTGGCGAAGGTGATCATCAACGACGAGTTCGGCATCTGCAAGGCGCTGATGACACAC
607 612 622 627
646 ATCCAGCGCTACACCCAGCGTCCAGAACTGCA
836 TCACCGACCTGACCGCGGAGCTGGCAAGTGGCCACCGTGACAGAGATCAACGCCGA
886 891
896 TGAAGCGTGGCGCGGAGGCGCGCTCAAGGGCATCTCAAGTACTAGTACGACCCCGATCG
943 948
956 TGTCCAGGACATGTCACCGATCCGATCCGACAGCTTGATCTTTCGACTGGGTCTGACCAAG
961 966
967 972
1016 TCATCGA
1040 TGTCTGTGTACGACAAGAGATGGGGCTATCTCAACCGCGCTGTGACCTGTGCGCCCTGG
1090 1095
1100 TCGGCAAGTCGTGTAGAGGGCGAGCGAAGCGAGCGGAGAAAGGCGCCATGCGCATCA
1107 1112 1127 1132 1144 1149 1157
1160 AGTCACTCGACGACGCTTCTCTCCGAAGGGGTGACGGGGCGGGGCTACTGTGCGCTCCG
1168 1173
1220 ACCTGAACGTGCCCTCTGAGGGGAGGACGATCAACCGACCGGGGCGCATCATGCGCTCGG

```

1225 1230      1240 1245      1270 1275
1280      TCGCAGCCTTGAGGCGCTTGAGTGAGCGCGCCAGCGTGCTGTCACCGCGCATCTGG
      1284 1289      1298 1303 1308      1321 1326
      1309 1314
1340      GCAGGCCCAAGGCTGAGCCGATCC
-----
6 matches found in sequence:
aaz11396 ; Extended DNA sequence for M. vaccae antigen GV-40P.
(from "mycobacterng.seq")
TOIG of: aaz11396 check: 5047 from: 1 to: 522

ID  AAZ11396 standard; DNA; 522 BP.
XX
AC  AAZ11396;
XX
DT  25-OCT-1999 (first entry)
XX
DE  Extended DNA sequence for M. vaccae antigen GV-40P.
XX
KW  Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW  dendritic cell maturation; infectious disease; immune disorder; cancer;
KW  respiratory system; mycobacterial infection; allergy; tuberculosis;
KW  leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW  dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW  squamous cell carcinoma; melanoma; ss.
XX
OS  Mycobacterium vaccae.
XX
PN  W09932634-AZ.
XX
PD  01-JUL-1999.
XX
PF  23-DEC-1998; 98WO-NZ00189.
XX
PR  04-DEC-1998; 98US-0205426.
PR  23-DEC-1997; 97US-0996624.
PR  23-DEC-1997; 97US-0997080.
PR  23-DEC-1997; 97US-0997362.
PR  11-JUN-1998; 98US-0095855.
PR  17-SEP-1998; 98US-0156181.
XX
PA  (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI  Prestidge RL, Skinner MA, Tan P, Vlaser ES, Watson J;
DR  WPI: 1999-430163/36.
XX
DR  P-PSDB; AAY14931.
XX
PT  Enhancing immune response to an antigen
XX
PS  Example 16; Page 242; 243pp; English.
XX
CC  The invention provides heat-killed Mycobacterium vaccae, or recombinant
CC  M. vaccae proteins. The M. vaccae proteins may be employed to activate
CC  T cells and natural killer cells, to stimulate the production of
CC  cytokines, to enhance the expression of co-stimulatory molecules on
CC  dendritic cells and monocytes, and to enhance dendritic cell maturation
CC  and function. The proteins can be expressed by standard recombinant
CC  methodology. Pharmaceutical compositions comprising the proteins or
CC  nucleic acid sequences encoding the proteins can be used for the
CC  treatment, prevention, and detection of disorders including infectious
CC  diseases, immune disorders and cancer. In particular, the compounds and
CC  methods are used for treatment of diseases of the respiratory system,
CC  such as mycobacterial infections, asthma, allergies, tuberculosis,
CC  leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
CC  psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
CC  alopecia areata, and skin cancers such as basal carcinoma, squamous cell
CC  carcinoma and melanoma.

```

```

XX  SQ      Sequence 522 BP; 102 A; 152 C; 182 G; 86 T; 0 other;
      AAZ11396 Length: 522 March 5, 2002 14:19 Type: N Check: 5047
      Found using 'seq2-3' (pappu403.key)
...
29      GCGCCGATCCCGCGCGAGTACATCCCTGCGTGATGCCGCGCGAGAGCCCATGC
      79      84
89      AGTACGGCGTGTGGCGCGCTACCCGCTGCTGTTACGTCAAGCTGACGCTGTCAGCGTG
      121 126      132 137
149      CCTACACAGAACTGCATCTGCTGGAAATGCAATTCACAG
...
220      GCCCAGCGCAGCCGCTGATCCTGGAGCCAGTATGCGCGTGCAGGTCAAGAGCGCCGAG
      270 275
280      GATTACATGGGTGAAGTAGCGCGACCTGAACCTCCCGCGTGTACAGATCCAGGCCANG
      301 306
340      GAGGAGCGGAGCGGTGC
...
457      GACTCTGACGCCGAAGTTCGCGCGAAGCTGTGAGAGAGATCATCGCGAAGCGAGCGC
      507 512
517      CAGTAA
-----
6 matches found in sequence:
aaz21131 ; Mycobacterium tuberculosis ESAT-6 nucleotide sequence.
(from "mycobacterng.seq")
TOIG of: aaz21131 check: 2938 from: 1 to: 591

ID  AAZ21131 standard; DNA; 591 BP.
XX
AC  AAZ21131;
XX
DT  18-NOV-1999 (first entry)
XX
DE  Mycobacterium tuberculosis ESAT-6 nucleotide sequence.
XX
KW  Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;
KW  delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;
KW  interferon-gamma release; ss.
XX
OS  Mycobacterium tuberculosis.
XX
FH  Key      Location/Qualifiers
FT  CDS      1..591
FT          /*tag= a
FT          /product= "ESAT-6"
XX
PN  W09945119-AZ.
XX
PD  10-SEP-1999.
XX
PF  05-MAR-1999; 99WO-DK00109.
XX
PR  06-MAR-1998; 98DK-0000306.
PR  06-MAR-1998; 98US-0077105.
XX
PA  (STAT-) STATENS SERUM INST.

```



```

XX  Jensen CL,  Folkersen J;
PI  MPI: 1999-551043/46.
XX  P-PSDB: AAY29888.
DR  New mycobacterial polypeptide produced in lactic acid bacteria, useful
XX  in tuberculosis diagnosis and vaccines
PS  Disclosure: Page 73-74; 76pp; English.
XX
CC  The present invention describes a bioactive polypeptide (or
CC  immunologically equivalent analogue) produced in lactic acid bacteria
CC  which reacts with lymphoid cells primed with Mycobacterium tuberculosis
CC  complex mycobacteria (M. tuberculosis, M. africanum or M. bovis). The
CC  polypeptide and ESAT-6 polypeptides are useful in compositions for
CC  diagnosis of and vaccination against tuberculosis caused by
CC  M. tuberculosis complex mycobacteria. The ESAT-6 polypeptide can be used
CC  to diagnose ongoing/previous sensitisation with these bacteria by
CC  detecting cytokine release when contacting blood samples with the
CC  polypeptide. The bioactive polypeptide may be used in diagnostic
CC  compositions and vaccines for mycobacteria other than of the
CC  M. tuberculosis complex, e.g. M. avium which infects poultry and
CC  occasionally humans, M. leprae; they are especially useful when they do
CC  not react with lymphoid cells previously primed with M. tuberculosis
CC  complex mycobacteria, and so do not give rise to a diagnostic reaction
CC  in individuals infected with these bacteria. The polypeptides may also
CC  be used in in vitro diagnostic tests e.g. stimulation of interferon-gamma
CC  release from lymphocytes. The polypeptide has similar or higher
CC  bioactivity as currently used tuberculin reagent in the standard
CC  delayed type hypersensitivity (DTH) skin test for tuberculosis, but may
CC  have greater specificity, being better able to discriminate between
CC  lymphoid cells primed from tuberculosis and from previous vaccination.
CC  The present sequence encodes M. tuberculosis ESAT-6 used in the
CC  exemplification of the present invention.
XX
SQ  Sequence 591 BP; 142 A; 164 C; 194 G; 91 T; 0 other;
AAZ21131 Length: 591 March 5, 2002 14:19 Type: N Check: 2938
Found using 'seq2-3' (pappu403.key)

```

```

aaz21132 : Mycobacterium tuberculosis ESAT-6 nucleotide sequence.
      (from "mycobacterium.seq")
TOIG of: aaz21132 check: 1165 from: 1 to: 288
ID  AAZ21132 standard; DNA: 288 BP.
XX
AC  AAZ21132;
XX
DT  18-NOV-1999 (first entry)
XX
DE  Mycobacterium tuberculosis ESAT-6 nucleotide sequence.
XX
KW  Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;
XX  delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;
XX  interferon-gamma release; ss.
XX
OS  Mycobacterium tuberculosis.
XX
FH  Key Location/Qualifiers
FT  1..288
FT  CDS /tag=a
FT  /product="ESAT-6"
XX
XX  WO945119-A2.
XX  10-SEP-1999.
XX
PF  05-MAR-1999; 99MO-DK00109.
XX
PR  06-MAR-1998; 98DK-0000306.
XX  06-MAR-1998; 98US-0077105.
XX
PA  (STAR-) STATENS SERUM INST.
XX
PI  Jensen CL, Folkersen J;
XX
DR  MPI: 1999-551043/46.
XX  P-PSDB: AAY29888.
XX
PT  New mycobacterial polypeptide produced in lactic acid bacteria, useful
XX  in tuberculosis diagnosis and vaccines
PS  Disclosure: Page 75; 76pp; English.
XX
CC  The present invention describes a bioactive polypeptide (or
CC  immunologically equivalent analogue) produced in lactic acid bacteria
CC  which reacts with lymphoid cells primed with Mycobacterium tuberculosis
CC  complex mycobacteria (M. tuberculosis, M. africanum or M. bovis). The
CC  polypeptide and ESAT-6 polypeptides are useful in compositions for
CC  diagnosis of and vaccination against tuberculosis caused by
CC  M. tuberculosis complex mycobacteria. The ESAT-6 polypeptide can be used
CC  to diagnose ongoing/previous sensitisation with these bacteria by
CC  detecting cytokine release when contacting blood samples with the
CC  polypeptide. The bioactive polypeptide may be used in diagnostic
CC  compositions and vaccines for mycobacteria other than of the
CC  M. tuberculosis complex, e.g. M. avium which infects poultry and
CC  occasionally humans, M. leprae; they are especially useful when they do
CC  not react with lymphoid cells previously primed with M. tuberculosis
CC  complex mycobacteria, and so do not give rise to a diagnostic reaction
CC  in individuals infected with these bacteria. The polypeptides may also
CC  be used in in vitro diagnostic tests e.g. stimulation of interferon-gamma
CC  release from lymphocytes. The polypeptide has similar or higher
CC  bioactivity as currently used tuberculin reagent in the standard
CC  delayed type hypersensitivity (DTH) skin test for tuberculosis, but may
CC  have greater specificity, being better able to discriminate between
CC  lymphoid cells primed from tuberculosis and from previous vaccination.
CC  The present sequence encodes M. tuberculosis ESAT-6 used in the
CC  exemplification of the present invention.
XX
SQ  Sequence 288 BP; 68 A; 80 C; 96 G; 44 T; 0 other;
AAZ21132 Length: 288 March 5, 2002 14:19 Type: N Check: 1165
Found using 'seq2-3' (pappu403.key)

```

3 matches found in sequence:

```

...
125 CCTGGGCGGTAGCGGTTGGAGGCGGTACAGGGGTGTCAGCAAAAATGGGAGCGCAGG      175 180
185 CTACCGAGCTGACACAGCGGCTGCAGAACCTGGCCGACCATTCAGCGCAAGCGGTCAGG      222 227
245 CAATGGCTTCGACCGAAGGCAACGCTCACTGGGATGTTGCATAG      265 270

9 matches found in sequence:
aa235746 : Human interleukin 17 receptor like protein encoding cDNA.
(from "mycobacterieng.seq")
TOIG of: aa235746 check: 7661 from: 1 to: 1816

ID AA235746 standard; cDNA: 1816 BP.
XX
AC AA235746;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human interleukin 17 receptor like protein encoding cDNA.
XX
KW Human; interleukin 17 receptor like protein; IL17R; IL-17;
KW diagnosis; detection; immune system related disorder; haemostasis;
KW cellular activation; angiogenesis; tumour metastasis; ovulation;
KW cellular migration; neurogenesis; infection; T-cell proliferation;
KW autoimmune disease; lymphocytic leukaemia; haematopoiesis; regulation;
KW sepsis; tumour; cancer; interstitial lung disease; arthritis;
KW lymphoma; immunosuppression; immunity; inflammatory bowel disease;
KW myelo suppression; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 10..1290
FT /tag= a
FT /product= "interleukin 17 receptor like protein"
FT sig_peptide 10..66
FT /tag= b
FT mat_peptide 67..1287
FT /tag= c
XX
XX WO914240-A1.
XX
XX 25-MAR-1999.
XX
XX 16-SEP-1998; 98WO-US19121.
XX
XX 17-SEP-1997; 97US-0059133.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Shi Y, Ruben SM;
XX
XX WPI: 2000-061918/05.
XX
XX P-PSDB: AAY49946.
XX
XX
XX New human interleukin-17 receptor like protein, e.g. to treat disorders
XX relating to cellular activation -
XX
XX Claim 2; Fig 1; 133p; English.
XX
XX The present sequence encodes human interleukin 17 receptor like protein
XX (IL17R), isolated from a cDNA library of human adult pulmonary tissue.
XX IL17R and its agonists can be used to treat disorders relating to
XX cellular activation, haemostasis, angiogenesis, tumour metastasis,

```

```

CC cellular migration and ovulation, and neurogenesis. They can also be
CC used to enhance host defences against resistant chronic and acute
CC infections, e.g. mycobacterial infections via the attraction and
CC activation of microbial leukocytes. IL17R may also be used to increase
CC T-cell proliferation by the stimulation of IL-2 biosynthesis for the
CC treatment of T-cell mediated autoimmune diseases and lymphocytic
CC leukaemias, to regulate haematopoiesis by regulating the activation and
CC differentiation of various haematopoietic progenitor cells, e.g. to
CC release mature leukocytes from the bone marrow following chemotherapy,
CC i.e. in stem cell mobilisation or to treat sepsis. The products can also
CC be used for the diagnosis or treatment of immune system related disorders
CC e.g. tumours, cancers, interstitial lung disease, and any dysregulation
CC of immune cell function including autoimmunity, arthritis, leukaemias,
CC lymphomas, immunosuppression, immunity, humoral immunity, inflammatory
CC bowel disease, or myelo suppression.
XX
XX Sequence 1816 BP; 532 A; 439 C; 399 G; 445 T; 1 other:
XX
AA235746 Length: 1816 March 5, 2002 14:18 Type: N Check: 7661 ..
Found using 'seq2-3' (pappu403.key)

1 GCACGAGGAGATGTCGCTGCTGCTAGACCTGGCCGCGCTGCGAGAGCGCCGTACCC      49 54
6 11 12 17
1 CGAGAGCCGACCGTTCAATGTGGCTGTGAACTGGGCCATCTCC
61
...
734 ATAGTGAAGGTGCTACGCTGAGCTGACCTCATATTTCTACTTGTGGCAGCGACTGCA      784 789
794 TCCGACATAAAGGAACAAGTGTGCTCTGCCCAACAACAGGCGCTTCCTTGATTA      832 837
854 ACAACAAAAGCAAGCCGGAGGCTGGCTGCTCT
...
1133 AGATGGGTCCAGTGCAGTGGCTTGCCACTCAAAAAGAGCAGACACAACAAAGTCGCTTCC      1183 1188
1193 TTCTTTCCATGAGAGTCAACAGAGTGTGCGATGCTGCTGTCGACAGCAGGCGAGTCC      1204 1209
1253 CCACTGA
...
1320 TTCAATCGACACAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAACAGCATTACA      1370 1375
1380 ATGCTCAGAGTGTGCCCCAAGTACCACTCATGAAGATGCCAC
...
1636 TTTTATACCAATAAATTTTCAATATTGCTAACTAATGTAGCATTAACGTACGATGGA      1686 1691
1696 AACTACATTTACAACTTCAAGCTGTTTATACATAGAAAATCAATT
...
3 matches found in sequence:

```

aa235748 ; Human interleukin 17 receptor like protein related cdna clone HETCC
(from "mycobacterieng.seq")
TOIG of: aa235748 check: 8080 from: 1 to: 327

ID AA235748 standard; cDNA; 327 BP.
XX
XX AA235748;
AC
XX
DT 01-FEB-2000 (first entry)
XX
DE Human interleukin 17 receptor like protein related cdna clone HETCC45RA.
XX
KW Human: interleukin 17 receptor like protein; IL17R; IL-17;
KW diagnosis: detection; immune system related disorder; haemostasis;
KW cellular activation; angiogenesis; tumour metastasis; ovulation;
KW cellular migration; neurogenesis; infection; T-cell proliferation;
KW autoimmune disease; lymphocytic leukaemia; haematopoiesis; regulation;
KW sepsis; tumour; cancer; interstitial lung disease; arthritis;
KW lymphoma; immunosuppression; immunity; inflammatory bowel disease;
KW myelo suppression; ss.
XX
OS Homo sapiens.
XX
XX WO914240-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
XX 16-SEP-1998; 98WO-US19121.
PF
XX
XX 17-SEP-1997; 97US-0059133.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Shi Y, Ruben SM;
PI
XX
XX WPI; 2000-061918/05.
DR
XX
XX New human interleukin-17 receptor like protein, e.g. to treat disorders
PT relating to cellular activation -
XX
XX
XX Claim 23; Page 128; 133pp; English.
PS

The present invention describes human interleukin 17 receptor like protein (IL17R), isolated from a cDNA library of human adult pulmonary tissue. The present sequence represents a human IL17R related cdna clone. IL17R and its agonists can be used to treat disorders relating to cellular activation, haemostasis, angiogenesis, tumour-metastasis, cellular migration and ovulation, and neurogenesis. They can also be used to enhance host defences against resistant chronic and acute infections, e.g. mycobacterial infections via the attraction and activation of microbial leukocytes. IL17R may also be used to increase T-cell proliferation by the stimulation of IL-2 biosynthesis for the treatment of T-cell mediated autoimmune diseases and lymphocytic leukaemias, to regulate haematopoiesis by regulating the activation and differentiation of various haematopoietic progenitor cells, e.g. to release mature leukocytes from the bone marrow following chemotherapy, i.e. in stem cell mobilisation or to treat sepsis. The products can also be used for the diagnosis or treatment of immune system related disorders e.g. tumours, cancers, interstitial lung disease, and any dysregulation of immune cell function including autoimmunity, arthritis, leukaemias, lymphomas, immunosuppression, immunity, humoral immunity, inflammatory bowel disease, or myelo suppression.

Sequence 327 BP; 69 A; 63 C; 86 G; 76 T; 33 other:
AA235748 Length: 327 March 5, 2002 14:18 Type: N Check: 8080
Found using 'seq2-3' (pappu403.key)

1 AATTCGGCANAGCCGGCGATGCGCTGCTGCTAGNCTGNGNCGCGTGTACAGAGC
16 21
22 27

61
63

1 match found in sequence:
aa235751; Human IL17R PCR 5' primer SEQ ID NO:8.
(from "mycobacterieng.seq")
TOIG of: aa235751 check: 3048 from: 1 to: 34

ID AA235751 standard; DNA; 34 BP.
XX
XX
AC AA235751;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human IL17R PCR 5' primer SEQ ID NO:8.
XX
XX
KW Human: interleukin 17 receptor like protein; IL17R; IL-17;
KW diagnosis: detection; immune system related disorder; haemostasis;
KW cellular activation; angiogenesis; tumour metastasis; ovulation;
KW cellular migration; neurogenesis; infection; T-cell proliferation;
KW autoimmune disease; lymphocytic leukaemia; haematopoiesis; regulation;
KW sepsis; tumour; cancer; interstitial lung disease; arthritis;
KW lymphoma; immunosuppression; immunity; inflammatory bowel disease;
KW myelo suppression; PCR primer; ss.
XX
XX
OS Synthetic.
XX
XX Homo sapiens.
XX
XX WO914240-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
XX 16-SEP-1998; 98WO-US19121.
PF
XX
XX 17-SEP-1997; 97US-0059133.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Shi Y, Ruben SM;
PI
XX
XX WPI; 2000-061918/05.
DR
XX
XX New human interleukin-17 receptor like protein, e.g. to treat disorders
PT relating to cellular activation -
XX
XX
XX Example 2; Page 95; 133pp; English.
PS

The present invention describes human interleukin 17 receptor like protein (IL17R), isolated from a cDNA library of human adult pulmonary tissue. The present sequence represents a PCR primer for human IL17R. IL17R and its agonists can be used to treat disorders relating to cellular activation, haemostasis, angiogenesis, tumour metastasis, cellular migration and ovulation, and neurogenesis. They can also be used to enhance host defences against resistant chronic and acute infections, e.g. mycobacterial infections via the attraction and activation of microbial leukocytes. IL17R may also be used to increase T-cell proliferation by the stimulation of IL-2 biosynthesis for the treatment of T-cell mediated autoimmune diseases and lymphocytic leukaemias, to regulate haematopoiesis by regulating the activation and differentiation of various haematopoietic progenitor cells, e.g. to release mature leukocytes from the bone marrow following chemotherapy, i.e. in stem cell mobilisation or to treat sepsis. The products can also be used for the diagnosis or treatment of immune system related disorders e.g. tumours, cancers, interstitial lung disease, and any dysregulation of immune cell function including autoimmunity, arthritis, leukaemias, lymphomas, immunosuppression, immunity, humoral immunity, inflammatory bowel disease, or myelo suppression.

CC Leukaemias, to regulate haematopoiesis by regulating the activation and
CC differentiation of various haematopoietic progenitor cells, e.g. to
CC release mature leukocytes from the bone marrow following chemotherapy,
CC i.e. in stem cell mobilisation or to treat sepsis. The products can also
CC be used for the diagnosis or treatment of immune system related disorders
CC e.g. tumours, cancers, interstitial lung disease, and any dysregulation
CC of immune cell function including autoimmunity, arthritis, leukaemias,
CC lymphomas, immunosuppression, immunity, humoral immunity, inflammatory
CC bowel disease, or myelo suppression.

XX Sequence 43 BP; 6 A; 16 C; 12 G; 9 T; 0 other;

AAZ3755 Length: 43 March 5, 2002 14:18 Type: N Check: 7995 ..
Found using 'seq2-3' (pappu403.key)

1 CTAGCCGATCCGACCATGTCGTGCTGCTGTAAGCCTGG
21 26

47 matches found in sequence:

aaz38383; M. tuberculosis genomic DNA comprising overlapping genes b2 and b3.
(from "mycobacterieng.seq")

TOIG of: aaz38383 check: 1829 from: 1 to: 2708

ID AAZ38383 standard: DNA; 2708 BP.

XX AAZ38383;

DT 22-FEB-2000 (first entry)

DE M. tuberculosis genomic DNA comprising overlapping genes b2 and b3.

XX Tuberculosis; overlapping gene: B2; B3; TMA; A60; antigenic;
KM thermostable macromolecular antigen complex; immunogenic; antigen;
KM antibody; species-specific; antiserum; recognition; epitope; diagnosis;
KM detection; immunosorbent assay; ELISA;
KM expression vector; vaccine; prevention; immune response; cellular; ss.

XX Mycobacterium tuberculosis.

OS Location/Qualifiers

XX Key

XX RBS

XX FT

XX FT

XX CDS

XX FT

XX FT

XX RBS

XX FT

XX FT

XX CDS

XX FT

XX FT

XX PN

XX EP962532-A1.

XX PD

XX 08-DEC-1999.

XX PF

XX 03-JUN-1998; 98EP-0870127.

XX PR

XX 03-JUN-1998; 98EP-0870127.

XX PA

XX (INNO-) INNOGENETICS NV.

XX PI

XX Cocito C, Coetsier C;

XX MPI: 2000-025637/03.

XX DR

XX P-PSDB; AAY52442, AAY52443.

XX PT

XX Mycobacteria tuberculosis polypeptides, useful for vaccines and in

PT vitro detection of tuberculosis in mammals

XX Claim 2; Fig 2; 46pp; English.

PS This sequence represents genomic DNA from Mycobacterium tuberculosis,
XX which comprises the overlapping genes b2 and b3, which respectively
CC encode proteins B2 (627 amino acids, AAY52442) and B3 (513 amino acids,
CC AAY52443). Proteins B2 and B3 form part of the thermostable
CC macromolecular antigen complex (TMA, also known as A60) of M.
CC tuberculosis. The protein components of A60 are responsible for its
CC immunogenicity; however, immunoassays using A60 as a reagent lack
CC species-specificity due to cross-reactions among homologous proteins of
CC different mycobacterial species. Proteins B2 and B3 are both recognised
CC by antisera from tuberculosis patients, and contain species-specific
CC epitopes (AAY52444, AAY52445). Proteins B2 and B3, or antigenic fragments
CC thereof may be used for the in vitro diagnosis of tuberculosis in a
CC mammal, especially a human, based on the detection of M. tuberculosis
CC antibodies via immunoassay (e.g., enzyme-linked immunosorbent assay,
CC ELISA). The proteins, fragments or expression vectors comprising DNA
CC encoding B2 and/or B3 may be used in a vaccine to prevent tuberculosis in
CC mammals. The proteins may also be used for detecting cellular immune
CC responses stimulated by tuberculosis, indicating contact and a protective
CC immune response, not disease. Prior art diagnostic procedures of the
CC Mycobacterium tuberculosis bacilli require very long incubation times,
CC and only mycobacteria with intact envelopes are identified by biopsies
CC using Ziehl-Neelsen staining. Currently available immunological
CC serological methods lack species-specificity. Use of new M. tuberculosis
CC antigens and selecting peptides which specifically react with sera from
CC infected subjects improves the specificity of immunological detection
CC assays.

XX Sequence 2708 BP; 535 A; 917 C; 804 G; 452 T; 0 other;

AAZ38383 Length: 2708 March 5, 2002 14:19 Type: N Check: 1829 ..
Found using 'seq2-3' (pappu403.key)

1 GACTGCAAGCTGGGTTGTCGCCGCGACACCGCATTCGCGCGGCGCCGATACG
44 49

61 AGATAGCGCAATGACCAACCCCTTCCACCTGTCCGACCG

...

126

TTGGGGGCAATGCCAGATCCACCTGCGCGGAGCTCGGGTTCACCGGAGTTGGCGG

186

TCAAGGTGCTGGCGGCTGATAGCCCGGATCCAGTTTACCTTCGCTCCGCGTG

246

AGCGCGAAGAGCGCGCGGATTCACACCTTCGATCGCGGCTACGACACCGGTC

306

AAGCCGAAGCGCGCGCGGATTCGCTACATTCGATGGAATACGTGACGAGGCGTA

366

CCCTGGCGACATTTGTCACACCGAAGGCGGATGACGCCAAGCGCGCATTCGAGTCA

426

TGCGCGAGCGCTGGCAAGCGCTGACTTCAATCATCAAGAACGATTCACCGTACG

486

TCAAGCGCGGCAATCATCATGATCAGCGCGGACCAATCATGAGTGAATGATTTGGCA

546

TGCGCGCGCGCATTTGCGGACAGCGGCAACAGCGTACCGACGAGCATGATCGCA

550
606 CGGGCAGTACCTGTCAACCCGACAGGCCCGGGTGATTCGGTCGAGCCCGATCCGATG
650 655
666 TCTATTCTTGGGCTGTGTCTTTATGAACTCTACCGGGGAGCCACCTTTTCACCGGCG
726 -1
ACTTACCCGCTCTCGGTGGCTTACCACATGTGCGGAAGCCGATCCACCTTCGGCGC
727
786 GGCACGAAAGCCTCTCCGCGACCTGGCGGTCTCTCAAGCGCTGGCCAAATAATC
812 817 829 834
846 CGGAAACCGCTATACAGACAGCGCGGAGATGCGCGCGACCTGTCGCGCTGCACACG
906 -1
GTGAGCCGCGCGAGCGCCCAAGTGTCTACCGATGCCGAGCGACCTCGCTGCTGT
919 924 961
966 -1
CTCGCGCGCGCAACTTACGGGTCCGCGACCGATCCGTAACAGCGCAGACTTAGACG
966
1026 -1
ACACCGACCGTGAACCGCAGCATCGGTTCGGTGGCGCTTGGGTTCGGTTCGCGCTGC
1027
1086 TCGCTGTGTGACCGGTGTGTGTAACCATCGCATACAGCTTCGGCGCATCATCCGCG
1112 1117
1146 -1
ACGTTCAAGTTCGCCGAGCTTCGGGGTCAATCTCCGCCGACCATGCCACACTGCAAA
1150 1160 1165 1184 1189 1190 1195
1206 ACCGGGGCTTCAAAATCCGCACCTTGACAGAGCCGAGCTC
1257 CGGACCAAGTTATCGGCACCGACCGCGCCGCAACAGCTGTGTGATGCGAGGAGAGAG
1307 1312
1317 TCACAGTCAACGTGTCCACCGGACCGGAGCAACGCAAAATACCGGAGTTCACCGTGA
1361 1366
1377 CATAGCCCGAAGCGGTCAAGAAACTGACTGCGCGGAGTT
1448 ACCCGGAAGTGGTGGGCAAGTCAATCGGAGACCAACCCGCGACCAACCGAGAGCTGGC
1498 1503
1508 ATACCAATGTGTATCATCATCGTTGGCTGTGTCGGCGGACGAAGACATTTCCGAT
1529 1534 1546 1551
1568 GTCCGGGGCCACAGCCTGACGTGGCGGAGAAAGAACCTCAACGCTTACGGCTTACACAAA
1607 1612
1628 TTCAGTCAAGCCTCGGTGAGACGCCCGCTCCCGC

1737 CCAAGGCAACCAATTCTGATGCCCCGACCTATCCGGCATGTTCTGGGTGAGCGGCAGAC
1787 1792
1797 CACGATTCGCGCGCTGGGCTGAGACGGGAGTCTGCAGAAAGGGGCGGACGACCGC
1844 1849 1850 1855
1857 GTGGCTCCCAACACAGCCGGGTCTGTATCAAAACCCGCGCGGAGACCGGCTTCAAC
1877 1882 1907 1912
1917 GGAAGGCGATCATCAGCTGAGGTTGGCCAGTACTTCCGCCGCG
1984 TCCGGTGAATGGCGGTGAGCAGCTTCTTTTCCAGCCGACGTAACAGGGGTGCTCTGC
2034 2039
2044 GTCCATCCGACAGAGTGAGCCAGTTGGCCAGTATCGGTGCCCGC
2156 CCATGATCACACCGCTGCTAGTGGCGCGGTGACCCCTTCAGACCGCTGGACGACTTAG
2206 2211
2216 GCAGAAATTTCAACGAATGATGAGTGCAGTGGCCGTGAAGAGATCCGG
2293 AATACGCTGTGTGTCTTGTCCGTGCAATAGCTCGGGGCGACGTCACAGGTGGCGCGAAC
2343 2348
2353 GCAAGCGCGATGGCTTGTGTGCCCAAGCGACGCCCCAGCAAAAGGGGTGTGTGCTCCG
2355 2360
2446 CGCTCCGGGGTACCGGAGCCGGGACTGAGCAGACACCGTTCGAATTGGCCGCGAGCGG
2496 2501
2506 GCCTCATCGATTAACCGGTGGTGTGCTGGCCACACACTTGGCTGATGCCAGCTGG
2526 2531 2529 2534
2566 CCGAGTACTGCAACCAAGTTGAACAGCAGAGTGTAGTTGTCAAGCAGCATCCGC
2610 2615
2626 ATCGTGACAGGCTACCGCTGACCCCTGCGGACGTCCTTCAGTATGTCAGAGGGGCTTGA
2659 2664
2686 GGCTGCCGGAAGTGCATCCGGAC
1 match found in sequence:
aa238384 : Mycobacterium tuberculosis h2 gene hybridisation probe TbbIot1.
(from "mycobacterieng.seq")
TOLG of: aa238384 check: 5200 from: 1 to: 20

ID AA238384 standard; DNA; 20 BP.
 XX
 AC AA238384;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis b2 gene hybridisation probe Tbb10t1.
 XX
 KW Tuberculosis; overlapping gene: B2; TWA; A60; antigenic;
 KW thermostable macromolecular antigen complex; immunogenic; antigen;
 KW antibody; species-specific; antiserum; recognition; epitope; diagnosis;
 KW detection; immunoassay; enzyme-linked immunosorbent assay; ELISA;
 KW expression vector; vaccine; prevention; immune response; cellular;
 KW hybridisation; probe; ss.
 XX
 OS Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT modified_base 1
 FT /tag= a
 FT /mod_base= OTHER
 FT /note= "5'-biotinylated"
 XX
 PN EP962532-A1.
 PD 08-DEC-1999.
 XX
 PF 03-JUN-1998; 98EP-0870127.
 XX
 PR 03-JUN-1998; 98EP-0870127.
 XX
 PA (INNO-) INNOGENETICS NV.
 PI Cocito C, Coetsier C;
 XX
 DR WPI; 2000-025637/03.
 XX
 PT Mycobacteria tuberculosis polypeptides, useful for vaccines and in
 PT vitro detection of tuberculosis in mammals -
 XX
 PS Example 1; Page 13; 46pp; English.
 XX
 CC This sequence represents a Mycobacterium tuberculosis b2 gene
 CC hybridisation probe Tbb10t1, used to identify upstream coding and
 CC regulatory sequences of the b2 gene from a phage library. The upstream
 CC DNA identified was then subcloned and sequenced. Protein B2 forms part
 CC of the thermostable macromolecular antigen complex (TWA, also known as
 CC A60) of M. tuberculosis. The protein components of A60 are responsible
 CC for its immunogenicity; however, immunoassays using A60 as a reagent
 CC lack species-specificity due to cross-reactions among homologous proteins
 CC of different mycobacterial species. Protein B2 is recognised by antisera
 CC from tuberculosis patients, and contains a species-specific epitope,
 CC K-14-R (AA92444). Protein B2, or antigenic fragments thereof may be
 CC used for the in vitro diagnosis of tuberculosis in a mammal, especially
 CC a human, based on the detection of M. tuberculosis antibodies via
 CC immunoassay (e.g., enzyme-linked immunosorbent assay, ELISA). The
 CC protein, fragments or expression vectors comprising DNA encoding B2
 CC may be used in a vaccine to prevent tuberculosis in mammals.
 CC The protein may also be used for detecting cellular immune responses
 CC stimulated by tuberculosis, indicating contact and a protective immune
 CC response, not disease. Prior art diagnostic procedures of the
 CC Mycobacterium tuberculosis bacilli require very long incubation times,
 CC and only mycobacteria with intact envelopes are identified by biopsies
 CC using Ziehl-Neelsen staining. Currently available immunoassays
 CC serological methods lack species-specificity. Use of new M. tuberculosis
 CC antigens and selecting peptides which specifically react with sera from
 CC infected subjects improves the specificity of immunological detection
 CC assays.
 CC
 XX Sequence 20 BP; 1 A; 7 C; 7 G; 5 T; 0 other;
 SQ
 AA238384 Length: 20 March 5, 2002 14:19 Type: N Check: 5200 ..

Found using 'seq2-3' (pappu403.key)

1
 4
 9
 GCGGCGCTTCGGCTTCACC

1 match found in sequence:

aa241856; IL-12 secretion inducing CpG oligonucleotide 1.
 (from "mycobacterieng.seq")
 TOIG of: aa241856 check: 8787 from: 1 to: 15

ID AA241856 standard; DNA; 15 BP.

XX
 AC AA241856;
 XX
 DT 24-JAN-2000 (first entry)
 XX
 DE IL-12 secretion inducing CpG oligonucleotide 1.
 XX

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX

OS Synthetic.

PN WO951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

PA (IOWA) UNIV IOWA RES FOUND.

PI Krieg AM, Weiner G;

DR WPI; 1999-620169/53.

PT Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX

PS Example 8; Page 67; 91pp; English.

XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
 XX which are used in the invention to induce interleukin-12 (IL-12)
 XX secretion from human PBMC. The invention comprises stimulating an immune
 XX response in a subject comprising administering to a subject exposed to an
 XX antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 XX oligonucleotide to induce a synergistic antigen specific immune
 XX response. The methods are useful for treating cancer by stimulating an
 XX antigen specific immune response against a cancer antigen. The methods
 XX can also be used to treat neoplastic disorders in humans, including but
 XX not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 XX neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 XX for treating infectious diseases, e.g. viral diseases such as HIV,
 XX bacterial diseases, and fungal diseases. The methods may also be used to
 XX treat allergic diseases, e.g. asthma. The methods and compositions may
 XX also be applied to treat cancer and tumours in non human subjects,
 XX e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 XX be treated and include leukæmia, haemangioepithelioma and bovine ocular
 XX neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 XX caused by the bacterium Corynebacterium pseudotuberculosis, and
 XX contagious lung tumour of sheep caused by jaagsiekte may also be
 XX treated. CpG oligonucleotides can be useful in activating B cells, NK
 XX cells, and antigen presenting cells, such as monocytes and macrophages.
 XX CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 XX can be used as an adjuvant in conjunction with tumour antigens to
 XX protect against a tumour challenge.

SQ Sequence 15 BP; 3 A; 3 C; 5 G; 4 T; 0 other;
 AA241856 Length: 15 March 5, 2002 14:19 Type: N Check: 8787
 Found using 'seq2-3' (pappu403.key)

1
 GCTAGACGTTAGCGT
 5 10

 1 match found in sequence:
 aaz41858 : IL-12 secretion inducing Cpg oligonucleotide 3.
 (from "mycobacterin_seg")
 TOIG of: aaz41858 check: 8787 from: 1 to: 15

ID AA241858 standard; DNA; 15 BP.
 XX AA241858;
 AC
 XX
 DT 24-JAN-2000 (first entry)
 DE IL-12 secretion inducing Cpg oligonucleotide 3.
 XX
 KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX
 OS Synthetic.
 XX
 PN WO951259-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 02-APR-1999; 99WO-US07335.
 XX
 PR 03-APR-1998; 98US-0080729.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 XX
 PI Krieg AM, Weiner G;
 XX
 DR WPI; 1999-620169/53.
 XX
 PT Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX
 PS
 XX

Example 8; Page 67; 91pp; English.

Sequences AA241856-241949 are phosphorothioate Cpg oligonucleotides
 which are used in the invention to induce interleukin-12 (IL-12)
 secretion from human PBMC. The invention comprises stimulating an immune
 response in a subject comprising administering to a subject exposed to an
 antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
 oligonucleotide to induce a synergistic antigen specific immune
 response. The methods are useful for treating cancer by stimulating an
 antigen specific immune response against a cancer antigen. The methods
 can also be used to treat neoplastic disorders in humans, including but
 not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 for treating infectious diseases, e.g. viral diseases such as HIV,
 bacterial diseases, and fungal diseases. The methods may also be used to
 treat allergic diseases, e.g. asthma. The methods and compositions may
 also be applied to treat cancer and tumours in non human subjects,
 e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 be treated and include leukaemia, haemangioepithelioma and bovine ocular
 neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 contagious lung tumour of sheep caused by jaagsiekte may also be
 treated. Cpg oligonucleotides can be useful in activating B cells, NK
 cells, and antigen presenting cells, such as monocytes and macrophages.
 Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and

CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX

SQ Sequence 15 BP; 3 A; 3 C; 5 G; 4 T; 0 other;

AA241858 Length: 15 March 5, 2002 14:19 Type: N Check: 8787
 Found using 'seq2-3' (pappu403.key)

1
 GCTAGACGTTAGCGT
 5 10

 1 match found in sequence:
 aaz41859 : IL-12 secretion inducing Cpg oligonucleotide 4.
 (from "mycobacterin_seg")
 TOIG of: aaz41859 check: 8787 from: 1 to: 15

ID AA241859 standard; DNA; 15 BP.
 XX AA241859;
 AC
 XX
 DT 24-JAN-2000 (first entry)
 DE IL-12 secretion inducing Cpg oligonucleotide 4.
 XX
 KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX
 OS Synthetic.
 XX
 PN WO951259-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 02-APR-1999; 99WO-US07335.
 XX
 PR 03-APR-1998; 98US-0080729.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 XX
 PI Krieg AM, Weiner G;
 XX
 DR WPI; 1999-620169/53.
 XX
 PT Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX
 PS
 XX

Example 8; Page 68; 91pp; English.

Sequences AA241856-241949 are phosphorothioate Cpg oligonucleotides
 which are used in the invention to induce interleukin-12 (IL-12)
 secretion from human PBMC. The invention comprises stimulating an immune
 response in a subject comprising administering to a subject exposed to an
 antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
 oligonucleotide to induce a synergistic antigen specific immune
 response. The methods are useful for treating cancer by stimulating an
 antigen specific immune response against a cancer antigen. The methods
 can also be used to treat neoplastic disorders in humans, including but
 not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 for treating infectious diseases, e.g. viral diseases such as HIV,
 bacterial diseases, and fungal diseases. The methods may also be used to
 treat allergic diseases, e.g. asthma. The methods and compositions may
 also be applied to treat cancer and tumours in non human subjects,
 e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 be treated and include leukaemia, haemangioepithelioma and bovine ocular
 neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 contagious lung tumour of sheep caused by jaagsiekte may also be

CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 15 BP; 3 A; 3 C; 5 G; 4 T; 0 other;

AAZ41859 Length: 15 March 5, 2002 14:19 Type: N Check: 8787 ..
 Found using 'seq2-3' (pappu403.key)

1 |-----|
 5 GCTAGCGTTAGCGT
 10

1 match found in sequence:
 aaz41860 : IL-12 secretion inducing Cpg oligonucleotide 5.
 (from "mycobacterin.seq")
 TOIG of: aaz41860 check: 8796 from: 1 to: 15

ID AAZ41860 standard; DNA: 15 BP.

AC AAZ41860;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing Cpg oligonucleotide 5.

KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

PI Krieg AM, Weiner G;

DR WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

PS Example 8; Page 68; 91pp; English.

CC Sequences AAZ41856-241949 are phosphorothioate Cpg oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukemia, haemangiopericytoma and bovine ocular

CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 15 BP; 3 A; 3 C; 5 G; 4 T; 0 other;

AAZ41860 Length: 15 March 5, 2002 14:19 Type: N Check: 8796 ..
 Found using 'seq2-3' (pappu403.key)

1 |-----|
 5 GCATGACGTTAGCT
 10

1 match found in sequence:
 aaz41861 : IL-12 secretion inducing Cpg oligonucleotide 6.
 (from "mycobacterin.seq")
 TOIG of: aaz41861 check: 5293 from: 1 to: 20

ID AAZ41861 standard; DNA: 20 BP.

AC AAZ41861;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing Cpg oligonucleotide 6.

KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

PI Krieg AM, Weiner G;

DR WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

PS Example 8; Page 68; 91pp; English.

CC Sequences AAZ41856-241949 are phosphorothioate Cpg oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may

CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

XX
SQ Sequence 20 BP; 4 A; 5 C; 6 G; 5 T; 0 other;

AAZ41861 Length: 20 March 5, 2002 14:19 Type: N Check: 5293 ..
Found using 'seq2-3' (pappu403.key)

1 ATGAGAGTCCAGCGTCTC
12 17

1 match found in sequence:
aaz41862 ; IL-12 secretion inducing Cpg oligonucleotide 7.
(from "mycobacterng.seq")
TOIG of: aaz41862 check: 5396 from: 1 to: 20

ID AAZ41862 standard; DNA; 20 BP.

XX
AC AAZ41862;

DT 24-JAN-2000 (first entry)

XX
DE IL-12 secretion inducing Cpg oligonucleotide 7.

KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.

XX
OS Synthetic.

XX
PN WO951259-A2.

XX
PD 14-OCT-1999.

XX
PF 02-APR-1999; 99WO-US07335.

XX
PR 03-APR-1998; 98US-0080729.

XX
PA (IOWA) UNIV IOWA RES FOUND.

XX
PI Krieg AM, Weiner G;

XX
DR WPI; 1999-620169/53.

XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -

XX
PS Example 8; Page 68; 91pp; English.

XX
CC Sequences AAZ41856-241949 are phosphorothioate Cpg oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful

CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

XX
SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

AAZ41862 Length: 20 March 5, 2002 14:19 Type: N Check: 5396 ..
Found using 'seq2-3' (pappu403.key)

1 ATGACTCTCGAGCGTCTC
12 17

1 match found in sequence:
aaz41863 ; IL-12 secretion inducing Cpg oligonucleotide 8.
(from "mycobacterng.seq")
TOIG of: aaz41863 check: 5396 from: 1 to: 20

ID AAZ41863 standard; DNA; 20 BP.

XX
AC AAZ41863;

DT 24-JAN-2000 (first entry)

XX
DE IL-12 secretion inducing Cpg oligonucleotide 8.

KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.

XX
OS Synthetic.

XX
PN WO951259-A2.

XX
PD 14-OCT-1999.

XX
PF 02-APR-1999; 99WO-US07335.

XX
PR 03-APR-1998; 98US-0080729.

XX
PA (IOWA) UNIV IOWA RES FOUND.

XX
PI Krieg AM, Weiner G;

XX
DR WPI; 1999-620169/53.

XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -

XX
PS Example 8; Page 69; 91pp; English.

XX
CC Sequences AAZ41856-241949 are phosphorothioate Cpg oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods

CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;
AAZ1863 Length: 20 March 5, 2002 14:19 Type: N Check: 5396 ..
Found using 'seq2-3' (pappu403.key)

1 ATGACTCTCGAGCGTTC 12 17
1-----1
1 match found in sequence:
aaz41864 ; IL-12 secretion inducing CpG oligonucleotide 9.
(from "mycobacteryn.seq")
TOIG of: aaz41864 check: 5396 from: 1 to: 20

ID AAZ41864 standard: DNA; 20 BP.
XX
XX AAZ41864;
AC
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 9.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KM neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KM antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
OS
XX
XX WO9951259-A2.
PN
XX
PD 14-OCT-1999.
XX
XX PF 02-APR-1999; 99WO-US07335.
XX
XX PR 03-APR-1998; 98US-0080729.
XX
XX PA (IOWA) UNTV IOWA RES FOUND.
XX
XX PI Krieg AM, Weiner G;
XX
XX DR WPI; 1999-620169/53.
XX
XX PT Novel synergistic combinations of immunostimulatory oligonucleotides
XX and immunopotentiating cytokines are useful for stimulating the immune
XX system ..
PS Example 8; Page 69; 91pp; English.
XX
XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG

CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;
AAZ1864 Length: 20 March 5, 2002 14:19 Type: N Check: 5396 ..
Found using 'seq2-3' (pappu403.key)

1 ATGACTCTCGAGCGTTC 12 17
1-----1
1 match found in sequence:
aaz41865 ; IL-12 secretion inducing CpG oligonucleotide 10.
(from "mycobacteryn.seq")
TOIG of: aaz41865 check: 5396 from: 1 to: 20

ID AAZ41865 standard: DNA; 20 BP.
XX
XX AAZ41865;
AC
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 10.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KM neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KM antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
OS
XX
XX WO9951259-A2.
PN
XX
PD 14-OCT-1999.
XX
XX PF 02-APR-1999; 99WO-US07335.
XX
XX PR 03-APR-1998; 98US-0080729.
XX
XX PA (IOWA) UNTV IOWA RES FOUND.
XX
XX PI Krieg AM, Weiner G;
XX
XX DR WPI; 1999-620169/53.
XX
XX PT Novel synergistic combinations of immunostimulatory oligonucleotides
XX and immunopotentiating cytokines are useful for stimulating the immune
XX system ..
PS Example 8; Page 69; 91pp; English.
XX
XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)

CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

AAZ41865 Length: 20 March 5, 2002 14:19 Type: N Check: 5396 ..
Found using 'seq2-3' (pappu403.key)

1 ATCGACTCTCGACGCTCTC
12 17

1 match found in sequence:
aaz41866 : IL-12 secretion inducing CpG oligonucleotide 11.
(from "mycobacterieng.seq")
TOIG of: aaz41866 check: 5318 from: 1 to: 20

ID AAZ41866 standard; DNA; 20 BP.

AC AAZ41866;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 11.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

PA (IOWA) UNIV IOWA RES FOUND.

PI Krieg AM, Weiner G;

XX MPI; 1999-620169/53.

PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX Example 8; Page 69; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

SQ Sequence 20 BP; 4 A; 7 C; 3 G; 6 T; 0 other;

AAZ41866 Length: 20 March 5, 2002 14:19 Type: N Check: 5318 ..
Found using 'seq2-3' (pappu403.key)

1 ATCGACTCTCGACGCTCTC
12 17

1 match found in sequence:
aaz41867 : IL-12 secretion inducing CpG oligonucleotide 12.
(from "mycobacterieng.seq")
TOIG of: aaz41867 check: 5134 from: 1 to: 20

ID AAZ41867 standard; DNA; 20 BP.

AC AAZ41867;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 12.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

PA (IOWA) UNIV IOWA RES FOUND.

PI Krieg AM, Weiner G;

XX MPI; 1999-620169/53.

PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX

```
PT system
XX
PS Example 8; Page 70; 91pp; English.
XX
CC Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases; e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 5 A; 6 C; 5 G; 4 T; 0 other;
AA241867 Length: 20 March 5, 2002 14:19 Type: N Check: 5134 ..
Found using 'seq2-3' (pappu403.key)

1
1 GAGAACGCTGCACCTTCAT
4 9
-----
1 match found in sequence:
aaz41868 ; IL-12 secretion inducing CpG oligonucleotide 13.
(from "mycobacterng.seq")
TOIG of: aaz41868 check: 5094 from: 1 to: 20

ID AA41868 standard; DNA; 20 BP.
XX
AC AA41868;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 13.
XX
KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
PN W09951259-A2.
XX
PD 14-OCT-1999.
XX
PE 02-APR-1999; 99WO-US07335.
XX
PR 03-APR-1998; 98US-0080729.
XX
PA (IOWA ) UNIV IOWA RES FOUND.
XX
PI Krieg AM, Weiner G;
XX
DR WPI; 1999-620169/53.
```

```
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
XX and immunopotentiating cytokines are useful for stimulating the immune
XX system
XX
PS Example 8; Page 70; 91pp; English.
XX
CC Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases; e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 5 A; 7 C; 4 G; 4 T; 0 other;
AA41868 Length: 20 March 5, 2002 14:19 Type: N Check: 5094 ..
Found using 'seq2-3' (pappu403.key)

1
1 GAGAACGCTGCACCTTCAT
4 9
-----
1 match found in sequence:
aaz41869 ; IL-12 secretion inducing CpG oligonucleotide 14.
(from "mycobacterng.seq")
TOIG of: aaz41869 check: 5166 from: 1 to: 20

ID AA41869 standard; DNA; 20 BP.
XX
AC AA41869;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 14.
XX
KW Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
PN W09951259-A2.
XX
PD 14-OCT-1999.
XX
PE 02-APR-1999; 99WO-US07335.
XX
PR 03-APR-1998; 98US-0080729.
XX
PA (IOWA ) UNIV IOWA RES FOUND.
XX
```

```
PI Krieg AM, Weiner G;
XX
XX WPI; 1999-620169/53.
DR
XX
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX Example 8; Page 70; 91pp; English.
PS
XX
XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 5 A; 6 C; 5 G; 4 T; 0 other;
AA241869 Length: 20 March 5, 2002 14:19 Type: N Check: 5166
Found using 'seq2-3' (pappu403.key)

1 |-----|
  4 9 GAGAACGCTCGACCTTCGAT

-----
1 match found in sequence:
aaz41872 : IL-12 secretion inducing CpG oligonucleotide 17.
(from "mycobacterng.seq")
TOIG of: aaz41872 check: 5134 from: 1 to: 20

ID AA241872 standard; DNA: 20 BP.
XX
XX AA241872;
AC
XX
XX 24-JAN-2000 (first entry)
DT
XX
XX IL-12 secretion inducing CpG oligonucleotide 17.
DE
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
OS
XX
XX W09951259-A2.
PN
XX
XX 14-OCT-1999.
PD
XX
XX 02-APR-1999; 99WO-US07335.
PF
XX
XX 03-APR-1998; 98US-0080729.
PR
```

```
XX
XX (IOWA ) UNIV IOWA RES FOUND.
PA
XX
XX Krieg AM, Weiner G;
PI
XX
XX WPI; 1999-620169/53.
DR
XX
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX Example 8; Page 71; 91pp; English.
PS
XX
XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 5 A; 6 C; 5 G; 4 T; 0 other;
AA241872 Length: 20 March 5, 2002 14:19 Type: N Check: 5134
Found using 'seq2-3' (pappu403.key)

1 |-----|
  4 9 GAGAACGCTCGACCTTCGAT

-----
1 match found in sequence:
aaz41873 : IL-12 secretion inducing CpG oligonucleotide 18.
(from "mycobacterng.seq")
TOIG of: aaz41873 check: 5118 from: 1 to: 20

ID AA241873 standard; DNA: 20 BP.
XX
XX AA241873;
AC
XX
XX 24-JAN-2000 (first entry)
DT
XX
XX IL-12 secretion inducing CpG oligonucleotide 18.
DE
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
OS
XX
XX W09951259-A2.
PN
XX
XX 14-OCT-1999.
PD
XX
```

```

PF 02-APR-1999; 99MO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
PR (IOWA ) UNIV IOWA RES FOUND.
XX
XX Kriegl AM, Weiner G;
XX
XX WPI; 1999-620169/53.
DR
XX
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX Example 8; Page 71; 91pp; English.
PS
XX
XX Sequences AA241856-241949 are phosphorothioate Cpg oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
XX Sequence 20 BP; 6 A; 5 C; 5 G; 4 T; 0 other;
SQ
AA241873 Length: 20 March 5, 2002 14:19 Type: N Check: 5118 ..
Found using 'seq2-3' (pappu403.key)

1
1 GAGACGATGACCTTCAT
4 9
-----
1 match found in sequence:
aaz41874 ; IL-12 secretion inducing Cpg oligonucleotide 19.
(from "mycobacterin.seq")
TOIG of: aaz41874 check: 4906 from: 1 to: 20

ID AA241874 standard; DNA: 20 BP.
XX
XX AA241874;
AC
XX
XX 24-JAN-2000 (first entry)
DT
XX
XX IL-12 secretion inducing Cpg oligonucleotide 19.
DE
XX
XX Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
OS
XX
XX W09951259-A2.
PN

```

```

XX
XX 14-OCT-1999.
PD
XX
XX 02-APR-1999; 99MO-US07335.
PF
XX
XX 03-APR-1998; 98US-0080729.
PR (IOWA ) UNIV IOWA RES FOUND.
XX
XX Kriegl AM, Weiner G;
XX
XX WPI; 1999-620169/53.
DR
XX
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX Example 8; Page 71; 91pp; English.
PS
XX
XX Sequences AA241856-241949 are phosphorothioate Cpg oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
XX Sequence 20 BP; 6 A; 6 C; 5 G; 3 T; 0 other;
SQ
AA241874 Length: 20 March 5, 2002 14:19 Type: N Check: 4906 ..
Found using 'seq2-3' (pappu403.key)

1
1 GAGACGCTCCAGACTGAT
4 9
-----
1 match found in sequence:
aaz41879 ; IL-12 secretion inducing Cpg oligonucleotide 24.
(from "mycobacterin.seq")
TOIG of: aaz41879 check: 5534 from: 1 to: 20

ID AA241879 standard; DNA: 20 BP.
XX
XX AA241879;
AC
XX
XX 24-JAN-2000 (first entry)
DT
XX
XX IL-12 secretion inducing Cpg oligonucleotide 24.
DE
XX
XX Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX

```

```
OS Synthetic.
XX
XX PN W09951259-A2.
XX
XX PD 14-OCT-1999.
XX
XX PF 02-APR-1999; 99WO-US07335.
XX
XX PR 03-APR-1998; 98US-0080729.
XX
XX PA (IOWA ) UNIV IOWA RES FOUND.
XX
XX PI Krieg AM, Weiner G;
XX
XX DR WPI: 1999-620169/53.
XX
XX PT Novel synergistic combinations of immunostimulatory oligonucleotides
XX PT and immunopotentiating cytokines are useful for stimulating the immune
XX PT system
XX
XX PS Example 8; Page 72; 91pp; English.
XX
XX PS Sequences AA241856-241949 are phosphorothioate Cpg oligonucleotides
XX CC which are used in the invention to induce interleukin-12 (IL-12)
XX CC secretion from human PBMC. The invention comprises stimulating an immune
XX CC response in a subject comprising administering to a subject exposed to an
XX CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
XX CC oligonucleotide to induce a synergistic antigen specific immune
XX CC response. The methods are useful for treating cancer by stimulating an
XX CC antigen specific immune response against a cancer antigen. The methods
XX CC can also be used to treat neoplastic disorders in humans, including but
XX CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
XX CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
XX CC for treating infectious diseases, e.g. viral diseases such as HIV,
XX CC bacterial diseases, and fungal diseases. The methods may also be used to
XX CC treat allergic diseases, e.g. asthma. The methods and compositions may
XX CC also be applied to treat cancer and tumours in non human subjects,
XX CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
XX CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
XX CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
XX CC caused by the bacterium Corynebacterium pseudotuberculosis, and
XX CC contagious lung tumour of sheep caused by jaagsiekte may also be
XX CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
XX CC cells, and antigen presenting cells, such as monocytes and macrophages.
XX CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
XX CC can be used as an adjuvant in conjunction with tumour antigens to
XX CC protect against a tumour challenge.
XX
XX SQ Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;
XX
XX AA241879 Length: 20 March 5, 2002 14:19 Type: N Check: 5534
XX Found using 'seq2-3' (pappu403.key)
XX
XX 1 |-----|
XX TCCATGACGTCCTCGATGCT
XX 6 11
XX
-----
1 match found in sequence:
aaz41881 ; IL-12 secretion inducing Cpg oligonucleotide 26.
(from "mycobacterieng.seq")
TOIG of: aaz41881 check: 5364 from: 1 to: 20
ID AA241881 standard; DNA: 20 BP.
XX
XX AC AA241881;
XX
XX DT 24-JAN-2000 (first entry)
XX
XX DE IL-12 secretion inducing Cpg oligonucleotide 26.
XX
XX KM Cpg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
XX human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
XX
```

```
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX OS Synthetic.
XX
XX OS W09951259-A2.
XX
XX PN 14-OCT-1999.
XX
XX PD 02-APR-1999; 99WO-US07335.
XX
XX PF 03-APR-1998; 98US-0080729.
XX
XX PR (IOWA ) UNIV IOWA RES FOUND.
XX
XX PA Krieg AM, Weiner G;
XX
XX PI WPI: 1999-620169/53.
XX
XX DR Novel synergistic combinations of immunostimulatory oligonucleotides
XX DR and immunopotentiating cytokines are useful for stimulating the immune
XX DR system
XX
XX PS Example 8; Page 74; 91pp; English.
XX
XX PS Sequences AA241856-241949 are phosphorothioate Cpg oligonucleotides
XX CC which are used in the invention to induce interleukin-12 (IL-12)
XX CC secretion from human PBMC. The invention comprises stimulating an immune
XX CC response in a subject comprising administering to a subject exposed to an
XX CC antigen, an immunopotentiating cytokine and an immunostimulatory Cpg
XX CC oligonucleotide to induce a synergistic antigen specific immune
XX CC response. The methods are useful for treating cancer by stimulating an
XX CC antigen specific immune response against a cancer antigen. The methods
XX CC can also be used to treat neoplastic disorders in humans, including but
XX CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
XX CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
XX CC for treating infectious diseases, e.g. viral diseases such as HIV,
XX CC bacterial diseases, and fungal diseases. The methods may also be used to
XX CC treat allergic diseases, e.g. asthma. The methods and compositions may
XX CC also be applied to treat cancer and tumours in non human subjects,
XX CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
XX CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
XX CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
XX CC caused by the bacterium Corynebacterium pseudotuberculosis, and
XX CC contagious lung tumour of sheep caused by jaagsiekte may also be
XX CC treated. Cpg oligonucleotides can be useful in activating B cells, NK
XX CC cells, and antigen presenting cells, such as monocytes and macrophages.
XX CC Cpg oligonucleotides enhance antibody dependent cellular cytotoxicity and
XX CC can be used as an adjuvant in conjunction with tumour antigens to
XX CC protect against a tumour challenge.
XX
XX SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;
XX
XX AA241881 Length: 20 March 5, 2002 14:19 Type: N Check: 5364
XX Found using 'seq2-3' (pappu403.key)
XX
XX 1 |-----|
XX ATGACCTCCGACGCTCTC
XX 12 17
XX
-----
1 match found in sequence:
aaz41882 ; IL-12 secretion inducing Cpg oligonucleotide 27.
(from "mycobacterieng.seq")
TOIG of: aaz41882 check: 5227 from: 1 to: 20
ID AA241882 standard; DNA: 20 BP.
XX
XX AC AA241882;
XX
XX DT 24-JAN-2000 (first entry)
XX
XX DE IL-12 secretion inducing Cpg oligonucleotide 27.
XX
```


XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX Synthetic.
 OS WO9951259-A2.
 PN 14-OCT-1999.
 XX
 XX 02-APR-1999; 99WO-US07335.
 PF
 XX 03-APR-1998; 98US-0080729.
 XX
 XX (IOWA) UNIV IOWA RES FOUND.
 PA
 PI Krieg AM, Weiner G;
 XX WPI; 1999-620169/53.
 DR
 XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX Example 8; Page 75; 91pp; English.
 PS
 XX Sequences AAZ41856-Z41949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX Sequence 20 BP; 5 A; 5 C; 5 G; 5 T; 0 other;
 SQ AAZ41882 Length: 20 March 5, 2002 14:19 Type: N Check: 5227
 Found using 'seq2-3' (pappu403.key)

1 ATAGGAGGTCACACGTTCTC
 12 17

 1 match found in sequence:
 aaz41883; IL-12 secretion inducing CpG oligonucleotide 28.
 (from "mycobacterng.seq")
 TOIG of: aaz41883 check: 5396 from: 1 to: 20

ID AAZ41883 standard; DNA: 20 BP.
 XX AAZ41883;
 AC
 XX

DT 24-JAN-2000 (first entry)
 XX
 DE IL-12 secretion inducing CpG oligonucleotide 28.
 XX
 XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX Synthetic.
 OS WO9951259-A2.
 PN 14-OCT-1999.
 XX
 XX 02-APR-1999; 99WO-US07335.
 PF
 XX 03-APR-1998; 98US-0080729.
 PR
 XX (IOWA) UNIV IOWA RES FOUND.
 PA
 PI Krieg AM, Weiner G;
 XX WPI; 1999-620169/53.
 DR
 XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX Example 8; Page 75; 91pp; English.
 PS
 XX Sequences AAZ41856-Z41949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;
 SQ AAZ41883 Length: 20 March 5, 2002 14:19 Type: N Check: 5396
 Found using 'seq2-3' (pappu403.key)

1 ATCGACTCTCGAGCGTTCTC
 12 17

 1 match found in sequence:
 aaz41884; IL-12 secretion inducing CpG oligonucleotide 29.
 (from "mycobacterng.seq")
 TOIG of: aaz41884 check: 5466 from: 1 to: 20
 ID AAZ41884 standard; DNA: 20 BP.

1 match found in sequence:
aaz41886 ; IL-12 secretion inducing CpG oligonucleotide 31.
(from "mycobacterng.seq")
TOIG of: aaz41886 check: 5396 from: 1 to: 20

ID AAZ41886 standard; DNA; 20 BP.
XX AAZ41886;
AC
XX
XX
XX
XX 24-JAN-2000 (first entry)
XX
XX IL-12 secretion inducing CpG oligonucleotide 31.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
XX
XX WO9951259-A2.
XX
XX 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Weiner G;
XX
XX WPI; 1999-620169/53.

Novel synergistic combinations of immunostimulatory oligonucleotides
and immunopotentiating cytokines are useful for stimulating the immune
system -

Example 8; Page 76; 91pp; English.

Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
which are used in the invention to induce interleukin-12 (IL-12)
secretion from human PBMC. The invention comprises stimulating an immune
response in a subject comprising administering to a subject exposed to an
antigen, an immunopotentiating cytokine and an immunostimulatory CpG
oligonucleotide to induce a synergistic antigen specific immune
response. The methods are useful for treating cancer by stimulating an
antigen specific immune response against a cancer antigen. The methods
can also be used to treat neoplastic disorders in humans, including but
not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
neuroblastoma, retinoblastoma, and glioma. The methods are also useful
for treating infectious diseases, e.g. viral diseases such as HIV,
bacterial diseases, and fungal diseases. The methods may also be used to
treat allergic diseases, e.g. asthma. The methods and compositions may
also be applied to treat cancer and tumours in non human subjects,
e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
be treated and include leukaemia, haemangioepithelioma and bovine ocular
neoplasia. Chronic, infectious, contagious diseases of sheep and goats
caused by the bacterium Corynebacterium pseudotuberculosis, and
contagious lung tumour of sheep caused by jaagsiekte may also be
treated. CpG oligonucleotides can be useful in activating B cells, NK
cells, and antigen presenting cells, such as monocytes and macrophages.
CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
can be used as an adjuvant in conjunction with tumour antigens to
protect against a tumour challenge.

Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

AAZ41886 Length: 20 March 5, 2002 14:19 Type: N Check: 5396
Found using 'seq2-3' (pappu403.key)

1 ATCGACTCTCGAGCGTTCCTC
|-----|

12 17

1 match found in sequence:
aaz41887 ; IL-12 secretion inducing CpG oligonucleotide 32.
(from "mycobacterng.seq")
TOIG of: aaz41887 check: 8796 from: 1 to: 15

ID AAZ41887 standard; DNA; 15 BP.
XX AAZ41887;
AC
XX
XX
XX 24-JAN-2000 (first entry)
XX
XX IL-12 secretion inducing CpG oligonucleotide 32.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
XX Synthetic.
XX
XX WO9951259-A2.
XX
XX 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Weiner G;
XX
XX WPI; 1999-620169/53.

Novel synergistic combinations of immunostimulatory oligonucleotides
and immunopotentiating cytokines are useful for stimulating the immune
system -

Example 8; Page 76; 91pp; English.

Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
which are used in the invention to induce interleukin-12 (IL-12)
secretion from human PBMC. The invention comprises stimulating an immune
response in a subject comprising administering to a subject exposed to an
antigen, an immunopotentiating cytokine and an immunostimulatory CpG
oligonucleotide to induce a synergistic antigen specific immune
response. The methods are useful for treating cancer by stimulating an
antigen specific immune response against a cancer antigen. The methods
can also be used to treat neoplastic disorders in humans, including but
not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
neuroblastoma, retinoblastoma, and glioma. The methods are also useful
for treating infectious diseases, e.g. viral diseases such as HIV,
bacterial diseases, and fungal diseases. The methods may also be used to
treat allergic diseases, e.g. asthma. The methods and compositions may
also be applied to treat cancer and tumours in non human subjects,
e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
be treated and include leukaemia, haemangioepithelioma and bovine ocular
neoplasia. Chronic, infectious, contagious diseases of sheep and goats
caused by the bacterium Corynebacterium pseudotuberculosis, and
contagious lung tumour of sheep caused by jaagsiekte may also be
treated. CpG oligonucleotides can be useful in activating B cells, NK
cells, and antigen presenting cells, such as monocytes and macrophages.
CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
can be used as an adjuvant in conjunction with tumour antigens to
protect against a tumour challenge.

Sequence 15 BP; 3 A; 3 C; 5 G; 4 T; 0 other;

AAZ41887 Length: 15 March 5, 2002 14:19 Type: N Check: 8796
Found using 'seq2-3' (pappu403.key)

CC protect against a tumour challenge.

XX Sequence 20 BP; 2 A; 6 C; 4 G; 8 T; 0 other;

SQ AAZ41894 Length: 20 March 5, 2002 14:19 Type: N Check: 5667 ..
Found using 'seq2-3' (pappu403.key)

1 TCCATGTCGTTCTCTGATGCT
6 11
|----|

1 match found in sequence:
aaz41895 ; IL-12 secretion inducing CpG oligonucleotide 40.
(from "mycobacterng.seq")
TOIG of: aaz41895 check: 5498 from: 1 to: 20

ID AAZ41895 standard; DNA; 20 BP.
XX AAZ41895;
AC
XX
XX 24-JAN-2000 (first entry)
DT
XX
DE IL-12 secretion inducing CpG oligonucleotide 40.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.

XX WO9951259-A2.
XX 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Weiner G;
XX WPI; 1999-620169/53.
XX
XX Novel synergistic combinations of immunostimulatory oligonucleotides
XX and immunopotentiating cytokines are useful for stimulating the immune
XX system -
XX
XX Example 8; Page 77; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
XX which are used in the invention to induce interleukin-12 (IL-12)
XX secretion from human PBMC. The invention comprises stimulating an immune
XX response in a subject comprising administering to a subject exposed to an
XX antigen, an immunopotentiating cytokine and an immunostimulatory CpG
XX oligonucleotide to induce a synergistic antigen specific immune
XX response. The methods are useful for treating cancer by stimulating an
XX antigen specific immune response against a cancer antigen. The methods
XX can also be used to treat neoplastic disorders in humans, including but
XX not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
XX neuroblastoma, retinoblastoma, and glioma. The methods are also useful
XX for treating infectious diseases, e.g. viral diseases such as HIV,
XX bacterial diseases, and fungal diseases. The methods may also be used to
XX treat allergic diseases, e.g. asthma. The methods and compositions may
XX also be applied to treat cancer and tumours in non human subjects,
XX e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
XX be treated and include leukaemia, haemangiopericytoma and bovine ocular
XX neoplasia. Chronic, infectious, contagious diseases of sheep and goats
XX caused by the bacterium Corynebacterium pseudotuberculosis, and
XX contagious lung tumour of sheep caused by jaagsiekte may also be
XX treated. CpG oligonucleotides can be useful in activating B cells, NK

CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

XX Sequence 20 BP; 4 A; 6 C; 3 G; 7 T; 0 other;
SQ AAZ41895 Length: 20 March 5, 2002 14:19 Type: N Check: 5498 ..
Found using 'seq2-3' (pappu403.key)

1 TCCATAACGTTCTCTGATGCT
6 11
|----|

1 match found in sequence:
aaz41896 ; IL-12 secretion inducing CpG oligonucleotide 41.
(from "mycobacterng.seq")
TOIG of: aaz41896 check: 5347 from: 1 to: 20

ID AAZ41896 standard; DNA; 20 BP.
XX AAZ41896;
AC
XX
XX 24-JAN-2000 (first entry)
DT
XX
DE IL-12 secretion inducing CpG oligonucleotide 41.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.

XX WO9951259-A2.
XX 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Weiner G;
XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
XX and immunopotentiating cytokines are useful for stimulating the immune
XX system -
XX
XX Example 8; Page 78; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
XX which are used in the invention to induce interleukin-12 (IL-12)
XX secretion from human PBMC. The invention comprises stimulating an immune
XX response in a subject comprising administering to a subject exposed to an
XX antigen, an immunopotentiating cytokine and an immunostimulatory CpG
XX oligonucleotide to induce a synergistic antigen specific immune
XX response. The methods are useful for treating cancer by stimulating an
XX antigen specific immune response against a cancer antigen. The methods
XX can also be used to treat neoplastic disorders in humans, including but
XX not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
XX neuroblastoma, retinoblastoma, and glioma. The methods are also useful
XX for treating infectious diseases, e.g. viral diseases such as HIV,
XX bacterial diseases, and fungal diseases. The methods may also be used to
XX treat allergic diseases, e.g. asthma. The methods and compositions may
XX also be applied to treat cancer and tumours in non human subjects,
XX e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
XX be treated and include leukaemia, haemangiopericytoma and bovine ocular
XX neoplasia. Chronic, infectious, contagious diseases of sheep and goats

CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

AAZ41896 Length: 20 March 5, 2002 14:19 Type: N Check: 5347

Found using 'seq2-3' (pappu403.key)

1 TCCATGACGTCCTCGATGCT
 6 11
 |-----|

 1 match found in sequence:
 aaz41898 ; IL-12 secretion inducing CpG oligonucleotide 43.
 (from "mycobacterng.seq")
 TOIG of: aaz41898 check: 3560 from: 1 to: 19

ID AAZ41898 standard; DNA; 19 BP.

AC AAZ41898;

XX 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 43.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

XX WO9951259-A2.

PN 14-OCT-1999.

XX 02-APR-1999; 99WO-US07335.

XX 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 XX and immunopotentiating cytokines are useful for stimulating the immune
 XX system

XX Example 8; Page 78; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,

CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

SQ Sequence 19 BP; 3 A; 3 C; 10 G; 3 T; 0 other;

AAZ41898 Length: 19 March 5, 2002 14:19 Type: N Check: 3560
 Found using 'seq2-3' (pappu403.key)

1 GGGGTCAACGTTGACGGGG
 |-----|
 7 12

 1 match found in sequence:
 aaz41900 ; IL-12 secretion inducing CpG oligonucleotide 45.
 (from "mycobacterng.seq")
 TOIG of: aaz41900 check: 9008 from: 1 to: 15

ID AAZ41900 standard; DNA; 15 BP.

XX AAZ41900;

XX 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 45.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

XX WO9951259-A2.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07335.

XX 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 XX and immunopotentiating cytokines are useful for stimulating the immune
 XX system

XX Example 8; Page 78; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,

CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX
 SQ Sequence 15 BP; 3 A; 2 C; 5 G; 5 T; 0 other;

AAZ41900 Length: 15 March 5, 2002 14:19 Type: N Check: 9008 ..
 Found using 'seq2-3' (pappu403.key)

1 GCTAGACGTTAGTGCT
 5 10
 |-----|

 1 match found in sequence:
 aaz41901 ; IL-12 secretion inducing CpG oligonucleotide 46.
 (from "mycobacterieng.seq")
 TOIG of: aaz41901 check: 9008 from: 1 to: 15

ID AAZ41901 standard; DNA; 15 BP.

XX AC AAZ41901;

XX DT 24-JAN-2000 (first entry)

XX DE IL-12 secretion inducing CpG oligonucleotide 46.

XX CPg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

XX OS Synthetic.

XX WO9951259-A2.

XX PD 14-OCT-1999.

XX PF 02-APR-1999; 99WO-US07335.

XX PR 03-APR-1998; 98US-0080729.

XX PA (IOWA) UNIV IOWA RES FOUND.

XX PI Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

XX Example 8; Page 79; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but

CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX SQ Sequence 15 BP; 3 A; 2 C; 5 G; 5 T; 0 other;

AAZ41901 Length: 15 March 5, 2002 14:19 Type: N Check: 9008 ..
 Found using 'seq2-3' (pappu403.key)

1 GCTAGACGTTAGTGCT
 5 10
 |-----|

1 match found in sequence:
 aaz41902 ; IL-12 secretion inducing CpG oligonucleotide 47.
 (from "mycobacterieng.seq")
 TOIG of: aaz41902 check: 7335 from: 1 to: 21

ID AAZ41902 standard; DNA; 21 BP.

XX AC AAZ41902;

XX DT 24-JAN-2000 (first entry)

XX DE IL-12 secretion inducing CpG oligonucleotide 47.

XX CPg oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

XX OS Synthetic.

XX WO9951259-A2.

XX PD 14-OCT-1999.

XX PF 02-APR-1999; 99WO-US07335.

XX PR 03-APR-1998; 98US-0080729.

XX PA (IOWA) UNIV IOWA RES FOUND.

XX PI Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

XX Example 8; Page 79; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune

CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX
 SQ Sequence 21 BP; 2 A; 6 C; 4 G; 9 T; 0 other;

AAZ41902 Length: 21 March 5, 2002 14:19 Type: N Check: 7335

Found using 'seq2-3' (pappu403.key)

```

1 1-----|
  6 11
  TCCATGTCGTTCTCTGATGCT

```

1 match found in sequence:

aa441903 ; IL-12 secretion inducing CpG oligonucleotide 48.

(from "mycobacterieng.seq")

TOIG of: aa41903 check: 5667 from: 1 to: 20

ID AAZ41903 standard; DNA; 20 BP.

AC AAZ41903;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 48.

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO9951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

PA Krieg AM, Weiner G;

PI WPI; 1999-620169/53.

DR Novel synergistic combinations of immunostimulatory oligonucleotides

PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system

XX Example 8; Page 79; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune

CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 20 BP; 2 A; 6 C; 4 G; 8 T; 0 other;

AAZ41903 Length: 20 March 5, 2002 14:19 Type: N Check: 5667

Found using 'seq2-3' (pappu403.key)

```

1 1-----|
  6 11
  TCCATGTCGTTCTCTGATGCT

```

1 match found in sequence:

aa441904 ; IL-12 secretion inducing CpG oligonucleotide 49.

(from "mycobacterieng.seq")

TOIG of: aa41904 check: 1989 from: 1 to: 24

ID AAZ41904 standard; DNA; 24 BP.

AC AAZ41904;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 49.

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO9951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

PA (IOWA) UNIV IOWA RES FOUND.

PI Krieg AM, Weiner G;

DR WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system

XX Example 8; Page 79; 91pp; English.

PA (IOWA) UNIV IOWA RES FOUND.
XX
PI Krieg AM, Weiner G;
XX
DR WPI; 1999-620169/53.
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
XX
PS Example 8; Page 82; 91pp; English.
XX
XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 1 A; 6 C; 4 G; 9 T; 0 other;
AAZ41919 Length: 20 March 5, 2002 14:19 Type: N Check: 6005 ..
Found using 'seq2-3' (pappu403.key)
1 TCCATGTCGTCCTGTCGTT
6 11 15 20

2 matches found in sequence:
aaz41920 ; IL-12 secretion inducing CpG oligonucleotide 65.
(from "mycobactering.seq")
TOIG of: aaz41920 check: 5580 from: 1 to: 20
ID AAZ41920 standard; DNA; 20 BP.
XX
AC AAZ41920;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 65.
XX
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
XX WO9951259-A2.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-US07335.

XX
PR 03-APR-1998; 98US-0080729.
XX
PA (IOWA) UNIV IOWA RES FOUND.
XX
PI Krieg AM, Weiner G;
XX
DR WPI; 1999-620169/53.
XX
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
XX
PS Example 8; Page 82; 91pp; English.
XX
XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;
AAZ41920 Length: 20 March 5, 2002 14:19 Type: N Check: 5580 ..
Found using 'seq2-3' (pappu403.key)
1 TCCATAGCGTCTCTAGCGTT
6 11 15 20

1 match found in sequence:
aaz41921 ; IL-12 secretion inducing CpG oligonucleotide 66.
(from "mycobactering.seq")
TOIG of: aaz41921 check: 7496 from: 1 to: 21
ID AAZ41921 standard; DNA; 21 BP.
XX
AC AAZ41921;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 66.
XX
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
XX WO9951259-A2.
XX
PN
XX

PD 14-OCT-1999.

XX 02-APR-1999; 99WO-US07335.

XX 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system

PS Example 8; Page 83; 91pp; English.

XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 21 BP; 0 A; 8 C; 4 G; 9 T; 0 other;

AA241921 Length: 21 March 5, 2002 14:19 Type: N Check: 7496
 Found using 'seq2-3' (pappu403.key)

```
1  |-----|
  3  TCGTCGCTGCTCCGCTTCTT
  8
```

 2 matches found in sequence:
 aaz41922 ; IL-12 secretion inducing CpG oligonucleotide 67.
 (from "mycobacterng.seq")
 TOIG of: aaz41922 check: 4122 from: 1 to: 19

ID AA241922 standard; DNA; 19 BP.

XX AA241922;

XX 24-JAN-2000 (first entry)

XX IL-12 secretion inducing CpG oligonucleotide 67.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX Synthetic.

OS

XX

PN WO9951259-A2.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07335.

XX 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides

PT and immunopotentiating cytokines are useful for stimulating the immune

PT system

XX Example 8; Page 83; 91pp; English.

XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 19 BP; 2 A; 6 C; 4 G; 7 T; 0 other;

AA241922 Length: 19 March 5, 2002 14:19 Type: N Check: 4122
 Found using 'seq2-3' (pappu403.key)

```
1  |-----| |-----|
  5  TCCTGACGTTCTCTGACGTT
  10 14 19
```

 2 matches found in sequence:
 aaz41923 ; IL-12 secretion inducing CpG oligonucleotide 68.
 (from "mycobacterng.seq")
 TOIG of: aaz41923 check: 4521 from: 1 to: 19

ID AA241923 standard; DNA; 19 BP.

XX AA241923;

XX 24-JAN-2000 (first entry)

XX IL-12 secretion inducing CpG oligonucleotide 68.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;

KW antigen presenting cell; infection; allergic disease.
 OS Synthetic.
 XX WO951259-A2.
 XX 14-OCT-1999.
 XX 02-APR-1999; 99WO-US07335.
 XX 03-APR-1998; 98US-0080729.
 XX (IOWA) UNIV IOWA RES FOUND.
 XX Krieg AM, Weiner G;
 XX WPI; 1999-620169/53.
 XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX Example 8; Page 83; 91pp; English.
 XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma, useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX SQ Sequence 19 BP; 0 A; 6 C; 4 G; 9 T; 0 other;
 AAZ41923 Length: 19 March 5, 2002 14:19 Type: N Check: 4521 ..
 Found using 'seq2-3' (pappu403.key)
 1 TCTGTCGTTCTCTGTCGTT
 5 10 14 19

 2 matches found in sequence:
 aaz41924 ; IL-12 secretion inducing CpG oligonucleotide 69.
 (from "mycobacterng.seq")
 TOIG of: aaz41924 check: 6430 from: 1 to: 20
 ID AAZ41924 standard; DNA; 20 BP.
 XX AAZ41924;
 AC AAZ41924;
 XX 24-JAN-2000 (first entry)
 DT IL-12 secretion inducing CpG oligonucleotide 69.
 XX

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX Synthetic.
 OS WO951259-A2.
 XX 14-OCT-1999.
 XX 02-APR-1999; 99WO-US07335.
 XX 03-APR-1998; 98US-0080729.
 XX (IOWA) UNIV IOWA RES FOUND.
 XX Krieg AM, Weiner G;
 XX WPI; 1999-620169/53.
 XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX Example 8; Page 83; 91pp; English.
 XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX SQ Sequence 20 BP; 1 A; 4 C; 4 G; 11 T; 0 other;
 AAZ41924 Length: 20 March 5, 2002 14:19 Type: N Check: 6430 ..
 Found using 'seq2-3' (pappu403.key)
 1 TCCATGTCGTTTGTGCGTT
 6 11 15 20

 2 matches found in sequence:
 aaz41925 ; IL-12 secretion inducing CpG oligonucleotide 70.
 (from "mycobacterng.seq")
 TOIG of: aaz41925 check: 6158 from: 1 to: 20
 ID AAZ41925 standard; DNA; 20 BP.
 XX AAZ41925;
 AC AAZ41925;
 XX 24-JAN-2000 (first entry)
 DT

```

XX DE IL-12 secretion inducing CpG oligonucleotide 70.
XX KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
XX KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
XX KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
XX KW antigen presenting cell; infection; allergic disease.
XX OS Synthetic.
XX PN WO9951259-A2.
XX PD 14-OCT-1999.
XX PF 02-APR-1999; 99WO-US07335.
XX PR 03-APR-1998; 98US-0080729.
XX PA (IOWA ) UNIV IOWA RES FOUND.
XX PI Krieg AM, Weiner G;
XX DR WPI; 1999-620169/53.
XX PT Novel synergistic combinations of immunostimulatory oligonucleotides
XX PT and immunopotentiating cytokines are useful for stimulating the immune
XX PT system.
XX PS Example 8; Page 83; 91pp; English.
XX CC Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
XX CC which are used in the invention to induce interleukin-12 (IL-12)
XX CC secretion from human PBMC. The invention comprises stimulating an immune
XX CC response in a subject comprising administering to a subject exposed to an
XX CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
XX CC oligonucleotide to induce a synergistic antigen specific immune
XX CC response. The methods are useful for treating cancer by stimulating an
XX CC antigen specific immune response against a cancer antigen. The methods
XX CC can also be used to treat neoplastic disorders in humans, including but
XX CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
XX CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
XX CC for treating infectious diseases, e.g. viral diseases such as HIV,
XX CC bacterial diseases, and fungal diseases. The methods may also be used
XX CC to treat allergic diseases, e.g. asthma. The methods and compositions may
XX CC also be applied to treat cancer and tumours in non human subjects,
XX CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
XX CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
XX CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
XX CC caused by the bacterium Corynebacterium pseudotuberculosis, and
XX CC contagious lung tumour of sheep caused by jaagsiekte may also be
XX CC treated. CpG oligonucleotides can be useful in activating B cells, NK
XX CC cells, and antigen presenting cells, such as monocytes and macrophages.
XX CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
XX CC can be used as an adjuvant in conjunction with tumour antigens to
XX CC protect against a tumour challenge.
XX SQ Sequence 20 BP; 0 A; 6 C; 4 G; 10 T; 0 other;

```

```

AAZ41925 Length: 20 March 5, 2002 14:19 Type: N Check: 6158
Found using 'seq2-3' (pappu403.key)

```

```

1 TCCCTGCTGCTCTCTGTCGTT
5 10 15 20

```

```

-----
2 matches found in sequence:
aaz41926 ; IL-12 secretion inducing CpG oligonucleotide 71.
(from "mycobacterng.seq")
TOIG of: aaz41926 check: 6081 from: 1 to: 20
ID AAZ41926 standard; DNA; 20 BP.
XX

```

```

AC AAZ41926;
XX DT 24-JAN-2000 (first entry)
XX DE IL-12 secretion inducing CpG oligonucleotide 71.
XX KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
XX KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
XX KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
XX KW antigen presenting cell; infection; allergic disease.
XX OS Synthetic.
XX PN WO9951259-A2.
XX PD 14-OCT-1999.
XX PF 02-APR-1999; 99WO-US07335.
XX PR 03-APR-1998; 98US-0080729.
XX PA (IOWA ) UNIV IOWA RES FOUND.
XX PI Krieg AM, Weiner G;
XX DR WPI; 1999-620169/53.
XX PT Novel synergistic combinations of immunostimulatory oligonucleotides
XX PT and immunopotentiating cytokines are useful for stimulating the immune
XX PT system.
XX PS Example 8; Page 84; 91pp; English.
XX CC Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
XX CC which are used in the invention to induce interleukin-12 (IL-12)
XX CC secretion from human PBMC. The invention comprises stimulating an immune
XX CC response in a subject comprising administering to a subject exposed to an
XX CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
XX CC oligonucleotide to induce a synergistic antigen specific immune
XX CC response. The methods are useful for treating cancer by stimulating an
XX CC antigen specific immune response against a cancer antigen. The methods
XX CC can also be used to treat neoplastic disorders in humans, including but
XX CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
XX CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
XX CC for treating infectious diseases, e.g. viral diseases such as HIV,
XX CC bacterial diseases, and fungal diseases. The methods may also be used
XX CC to treat allergic diseases, e.g. asthma. The methods and compositions may
XX CC also be applied to treat cancer and tumours in non human subjects,
XX CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
XX CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
XX CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
XX CC caused by the bacterium Corynebacterium pseudotuberculosis, and
XX CC contagious lung tumour of sheep caused by jaagsiekte may also be
XX CC treated. CpG oligonucleotides can be useful in activating B cells, NK
XX CC cells, and antigen presenting cells, such as monocytes and macrophages.
XX CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
XX CC can be used as an adjuvant in conjunction with tumour antigens to
XX CC protect against a tumour challenge.
XX SQ Sequence 20 BP; 0 A; 6 C; 4 G; 10 T; 0 other;

```

```

AAZ41926 Length: 20 March 5, 2002 14:19 Type: N Check: 6081
Found using 'seq2-3' (pappu403.key)

```

```

1 TCCCTGCTGCTCTCTGTCGTT
6 11 15 20

```

```

-----
2 matches found in sequence:
aaz41927 ; IL-12 secretion inducing CpG oligonucleotide 72.
(from "mycobacterng.seq")
TOIG of: aaz41927 check: 6549 from: 1 to: 20

```

ID AAZ41927 standard; DNA; 20 BP.
AC AAZ41927;
DT 24-JAN-2000 (first entry)
XX IL-12 secretion inducing CpG oligonucleotide 72.
DE
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX Synthetic.
OS WO9951259-A2.
PN 14-OCT-1999.
PD
XX 02-APR-1999; 99WO-US07335.
XX
PF 03-APR-1998; 98US-0080729.
PR
XX (IOWA) UNIV IOWA RES FOUND.
PA
XX Krieg AM, Weiner G;
PI WPI; 1999-620169/53.
DR
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
XX Example 8; Page 84; 91pp; English.
PS
XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 0 A; 4 C; 4 G; 12 T; 0 other;

AAZ41927 Length: 20 March 5, 2002 14:19 Type: N Check: 6549
Found using 'seq2-3' (pappu403.key)

1 TCCTGTCGTTTTCGCTT
5 10 15 20
|-----| |-----|

1 match found in sequence:

aaZ41928 ; IL-12 secretion inducing CpG oligonucleotide 73.
(from "mycobacterng.seq")
TOIG of: aaZ41928 check: 7488 from: 1 to: 21

ID AAZ41928 standard; DNA; 21 BP.
XX
AC AAZ41928;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 73.
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX Synthetic.
OS
XX WO9951259-A2.
PN
XX 14-OCT-1999.
PD
XX 02-APR-1999; 99WO-US07335.
XX
PF 03-APR-1998; 98US-0080729.
PR
XX (IOWA) UNIV IOWA RES FOUND.
PA
XX Krieg AM, Weiner G;
PI WPI; 1999-620169/53.
DR
XX Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
XX Example 8; Page 84; 91pp; English.

Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
which are used in the invention to induce interleukin-12 (IL-12)
secretion from human PBMC. The invention comprises stimulating an immune
response in a subject comprising administering to a subject exposed to an
antigen, an immunopotentiating cytokine and an immunostimulatory CpG
oligonucleotide to induce a synergistic antigen specific immune
response. The methods are useful for treating cancer by stimulating an
antigen specific immune response against a cancer antigen. The methods
can also be used to treat neoplastic disorders in humans, including but
not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
neuroblastoma, retinoblastoma, and glioma. The methods are also useful
for treating infectious diseases, e.g. viral diseases such as HIV,
bacterial diseases, and fungal diseases. The methods may also be used to
treat allergic diseases, e.g. asthma. The methods and compositions may
also be applied to treat cancer and tumours in non human subjects,
e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
be treated and include leukaemia, haemangioepithelioma and bovine ocular
neoplasia. Chronic, infectious, contagious diseases of sheep and goats
caused by the bacterium Corynebacterium pseudotuberculosis, and
contagious lung tumour of sheep caused by jaagsiekte may also be
treated. CpG oligonucleotides can be useful in activating B cells, NK
cells, and antigen presenting cells, such as monocytes and macrophages.
CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
can be used as an adjuvant in conjunction with tumour antigens to
protect against a tumour challenge.

Sequence 21 BP; 0 A; 8 C; 4 G; 9 T; 0 other;

AAZ41928 Length: 21 March 5, 2002 14:19 Type: N Check: 7488
Found using 'seq2-3' (pappu403.key)

1 |-----|
TCGTGCTGCTGCTGCTCTCTT
3 8

```
-----
1 2 matches found in sequence:
  aaz41929 : IL-12 secretion inducing CpG oligonucleotide 74.
  (from "mycobacterng.seq")
  TOIG of: aaz41929 check: 8020 from: 1 to: 21

ID AAZ41929 standard; DNA; 21 BP.
XX
AC AAZ41929;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 74.
XX
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
PN WO9951259-A2.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-US07335.
XX
PR 03-APR-1998; 98US-0080729.
XX
PA (IOWA ) UNIV IOWA RES FOUND.
XX
PI Krieg AM, Weiner G;
XX
DR WPI; 1999-620169/53.
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
XX
PS Example 8; Page 84; 91pp; English.
XX
CC Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods are also useful
CC to treat allergic diseases, e.g. asthma. The methods may also be used to
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 21 BP; 0 A; 5 C; 5 G; 11 T; 0 other;

AAZ41929 Length: 21 March 5, 2002 14:19 Type: N Check: 8020
Found using 'seq2-3' (pappu403.key)

-----
1 2 matches found in sequence:
  aaz41930 : IL-12 secretion inducing CpG oligonucleotide 75.
  (from "mycobacterng.seq")
  TOIG of: aaz41930 check: 6005 from: 1 to: 20

ID AAZ41930 standard; DNA; 20 BP.
XX
AC AAZ41930;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 75.
XX
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
PN WO9951259-A2.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-US07335.
XX
PR 03-APR-1998; 98US-0080729.
XX
PA (IOWA ) UNIV IOWA RES FOUND.
XX
PI Krieg AM, Weiner G;
XX
DR WPI; 1999-620169/53.
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
XX
PS Example 8; Page 85; 91pp; English.
XX
CC Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods are also useful
CC to treat allergic diseases, e.g. asthma. The methods may also be used to
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 1 A; 6 C; 4 G; 9 T; 0 other;
```


AAZ41930 Length: 20 March 5, 2002 14:19 Type: N Check: 6005 ..
Found using 'seq2-3' (pappu403.key)

1 TCCATGTCGTTCGTCGTT
6 11 15 20
|----| |----|
6 11 15 20

2 matches found in sequence:
aaz41934 ; IL-12 secretion inducing CpG oligonucleotide 79.
(from "mycobacterng.seq")
TOIG of: aaz41934 check: 5670 from: 1 to: 20

ID AAZ41934 standard; DNA; 20 BP.
XX AAZ41934;
AC
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 79.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
PN WO9951259-A2.
XX
PD 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Welner G;
XX
XX WPI; 1999-620169/53.
XX

Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
PS Example 8; Page 85; 91pp; English.
XX

Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.

XX Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;
SQ
AAZ41934 Length: 20 March 5, 2002 14:19 Type: N Check: 5670 ..
Found using 'seq2-3' (pappu403.key)

1 TCCACGACGTTTCGACGTT
6 11 15 20
|----| |----|
6 11 15 20

3 matches found in sequence:
aaz41935 ; IL-12 secretion inducing CpG oligonucleotide 80.
(from "mycobacterng.seq")
TOIG of: aaz41935 check: 6226 from: 1 to: 20

ID AAZ41935 standard; DNA; 20 BP.
XX
AC AAZ41935;
XX
DT 24-JAN-2000 (first entry)
XX
DE IL-12 secretion inducing CpG oligonucleotide 80.
XX
XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.
XX
PN WO9951259-A2.
XX
PD 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Welner G;
XX
XX WPI; 1999-620169/53.
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system -
PS Example 8; Page 86; 91pp; English.
XX

Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.

CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

SQ Sequence 20 BP; 0 A; 4 C; 6 G; 10 T; 0 other;

AAZ41935 Length: 20 March 5, 2002 14:20 Type: N Check: 6226
 Found using 'seq2-3' (pappu403.key)

```

-----
1      |-----| |-----| |-----|
      3      8      15      20
      TCGTCGTTGTCGTTGTCGTT
      3      8      9      14

```

3 matches found in sequence:
 aaz41936 ; IL-12 secretion inducing CpG oligonucleotide 81.
 (from "mycobacterng.seq")
 TOIG of: aaz41936 check: 3528 from: 1 to: 24

ID AAZ41936 standard; DNA: 24 BP.

AC AAZ41936;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 81.

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO9951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

XX Example 8; Page 86; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular

CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

SQ Sequence 24 BP; 0 A; 4 C; 6 G; 14 T; 0 other;

AAZ41936 Length: 24 March 5, 2002 14:20 Type: N Check: 3528
 Found using 'seq2-3' (pappu403.key)

```

-----
1      |-----| |-----| |-----| |-----|
      3      8      11      16      19      24
      TCGTCGTTGTCGTTGTCGTT

```

3 matches found in sequence:
 aaz41937 ; IL-12 secretion inducing CpG oligonucleotide 82.
 (from "mycobacterng.seq")
 TOIG of: aaz41937 check: 9752 from: 1 to: 22

ID AAZ41937 standard; DNA: 22 BP.

AC AAZ41937;

DT 24-JAN-2000 (first entry)

DE IL-12 secretion inducing CpG oligonucleotide 82.

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.

PN WO9951259-A2.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-US07335.

PR 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

XX Example 8; Page 86; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular

CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 22 BP; 0 A; 4 C; 6 G; 12 T; 0 other;

AAZ41937 Length: 22 March 5, 2002 14:20 Type: N Check: 9752
 Found using 'seq2-3' (pappu403.key)

```

1  TCGTCGTTGCGTTTTCGTTGCGTT  |-----| |-----|
    3      8      9      14      17      22

```

 2 matches found in sequence:
 aaz41938 ; IL-12 secretion inducing CpG oligonucleotide 83.
 (from "mycobacterieng.seq")
 TOIG of: aaz41938 check: 7796 from: 1 to: 21

ID AAZ41938 standard; DNA; 21 BP.
 XX
 AC AAZ41938;
 XX
 DT 24-JAN-2000 (first entry)
 XX
 DE IL-12 secretion inducing CpG oligonucleotide 83.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

XX Synthetic.

XX WO9951259-A2.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07335.

XX 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

XX Example 8; Page 86; 91pp; English.

XX Sequences AAZ41856-Z41949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but

CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangioepithelioma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX Sequence 21 BP; 0 A; 4 C; 8 G; 9 T; 0 other;

AAZ41938 Length: 21 March 5, 2002 14:20 Type: N Check: 7796
 Found using 'seq2-3' (pappu403.key)

```

1  GCGTCGCTGTCGTTGTCGTT  |-----| |-----|
    10      15      16      21

```

 3 matches found in sequence:
 aaz41940 ; IL-12 secretion inducing CpG oligonucleotide 85.
 (from "mycobacterieng.seq")
 TOIG of: aaz41940 check: 8024 from: 1 to: 21

ID AAZ41940 standard; DNA; 21 BP.

XX AAZ41940;

XX 24-JAN-2000 (first entry)

XX IL-12 secretion inducing CpG oligonucleotide 85.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

XX Synthetic.

XX WO9951259-A2.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07335.

XX 03-APR-1998; 98US-0080729.

XX (IOWA) UNIV IOWA RES FOUND.

XX Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -

XX Example 8; Page 87; 91pp; English.

XX Sequences AAZ41856-Z41949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an

CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by *jaagsiekte* may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX
 SQ Sequence 21 BP; 0 A; 3 C; 6 G; 12 T; 0 other;

AAZ41940 Length: 21 March 5, 2002 14:20 Type: N Check: 8024

Found using 'seq2-3' (pappu403.key)

```

1      |-----| |-----| |-----|
      TGTGCGTTGTCGTTGTCGTT
        2   7   9   14   21

```

 4 matches found in sequence:
 aaz41941 ; IL-12 secretion inducing CpG oligonucleotide 86.
 (from "mycobacterieng.seq")
 TOIG of: aaz41941 check: 5116 from: 1 to: 25

ID AAZ41941 standard; DNA; 25 BP.
 XX
 AC AAZ41941;
 XX
 DT 24-JAN-2000 (first entry)
 XX
 DE IL-12 secretion inducing CpG oligonucleotide 86.
 XX
 KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.
 XX
 XX WO9951259-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 02-APR-1999; 99WO-US07335.
 XX
 PR 03-APR-1998; 98US-0080729.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 XX
 XX Krieg AM, Weiner G;
 XX
 XX WPI; 1999-620169/53.
 DR
 XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system

PS Example 8; Page 87; 91pp; English.
 XX
 CC Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides

CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium *Corynebacterium pseudotuberculosis*, and
 CC contagious lung tumour of sheep caused by *jaagsiekte* may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX
 SQ Sequence 25 BP; 0 A; 4 C; 8 G; 13 T; 0 other;

AAZ41941 Length: 25 March 5, 2002 14:20 Type: N Check: 5116
 Found using 'seq2-3' (pappu403.key)

```

1      |-----| |-----| |-----| |-----|
      TGTGCGTTGTCGTTGTCGTTGTCGTT
        2   7   8   13   14   19   20   25

```

 3 matches found in sequence:
 aaz41942 ; IL-12 secretion inducing CpG oligonucleotide 87.
 (from "mycobacterieng.seq")
 TOIG of: aaz41942 check: 4709 from: 1 to: 19

ID AAZ41942 standard; DNA; 19 BP.
 XX
 AC AAZ41942;
 XX
 DT 24-JAN-2000 (first entry)
 XX
 DE IL-12 secretion inducing CpG oligonucleotide 87.
 XX
 KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

OS Synthetic.
 XX
 XX WO9951259-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 02-APR-1999; 99WO-US07335.
 XX
 PR 03-APR-1998; 98US-0080729.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 XX
 XX Krieg AM, Weiner G;
 XX
 XX WPI; 1999-620169/53.
 DR

XX Novel synergistic combinations of immunostimulatory oligonucleotides
 PT and immunopotentiating cytokines are useful for stimulating the immune

PT system -
 XX Example 8; Page 87; 91pp; English.
 PS
 CC Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX SQ Sequence 19 BP; 0 A; 3 C; 6 G; 10 T; 0 other;

AAZ41942 Length: 19 March 5, 2002 14:20 Type: N Check: 4709
 Found using 'seq2-3' (pappu403.key)

```

1  |-----| |-----|
    |-----| |-----|
    2  TCGTCGTCGTCGTCGTT
      7  14  19
      8  13

```

3 matches found in sequence:
 aaz41943 ; IL-12 secretion inducing CpG oligonucleotide 88.
 (from "mycobacterng.seq")
 TOIG of: aaz41943 check: 7988 from: 1 to: 14

ID AAZ41943 standard; DNA; 14 BP.

XX AC AAZ41943;

XX DT 24-JAN-2000 (first entry)

XX DE IL-12 secretion inducing CpG oligonucleotide 88.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

XX OS Synthetic.

XX PN WO9951259-A2.

XX PD 14-OCT-1999.

XX PF 02-APR-1999; 99WO-US07335.

XX PR 03-APR-1998; 98US-0080729.

XX PA (IOWA) UNIV IOWA RES FOUND.

XX PI Krieg AM, Weiner G;

XX WPI; 1999-620169/53.

XX Novel synergistic combinations of immunostimulatory oligonucleotides

PT and immunopotentiating cytokines are useful for stimulating the immune

PT system -

XX Example 8; Page 87; 91pp; English.

XX Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 CC which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.

XX SQ Sequence 14 BP; 0 A; 4 C; 4 G; 6 T; 0 other;

AAZ41943 Length: 14 March 5, 2002 14:20 Type: N Check: 7988
 Found using 'seq2-3' (pappu403.key)

```

1  |-----| |-----|
    |-----| |-----|
    3  TCGTCGTCGTCGTT
      6  11  14

```

2 matches found in sequence:

aaz41944 ; IL-12 secretion inducing CpG oligonucleotide 89.

(from "mycobacterng.seq")

TOIG of: aaz41944 check: 7068 from: 1 to: 13

ID AAZ41944 standard; DNA; 13 BP.

XX AC AAZ41944;

XX DT 24-JAN-2000 (first entry)

XX DE IL-12 secretion inducing CpG oligonucleotide 89.

XX CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.

XX OS Synthetic.

XX PN WO9951259-A2.

XX PD 14-OCT-1999.

XX

```

PF 02-APR-1999; 99WO-US07335.
XX
PR 03-APR-1998; 98US-0080729.
XX
PA (IOWA ) UNIV IOWA RES FOUND.
XX
PI Krieg AM, Weiner G;
XX
XX WPI; 1999-620169/53.
DR
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
PS
PS Example 8; Page 87; 91pp; English.
XX
XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 13 BP; 0 A; 2 C; 4 G; 7 T; 0 other;

AA241944 Length: 13 March 5, 2002 14:20 Type: N Check: 7068
Found using 'seq2-3' (pappu403.key)

```

```

1 TGTCGTTGTCGTT
2 7
8 13
|-----| |-----|
|-----| |-----|

2 matches found in sequence:
aaz41945 ; IL-12 secretion inducing CpG oligonucleotide 90.
(from "mycobacterng.seq")
TOIG of: aaz41945 check: 5580 from: 1 to: 20

ID AA241945 standard; DNA; 20 BP.
XX
AC AA241945;
XX
XX 24-JAN-2000 (first entry)
XX
XX IL-12 secretion inducing CpG oligonucleotide 90.
XX
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
KW antigen presenting cell; infection; allergic disease.
XX
OS Synthetic.

```

```

XX WO9951259-A2.
XX
XX 14-OCT-1999.
XX
XX 02-APR-1999; 99WO-US07335.
XX
XX 03-APR-1998; 98US-0080729.
XX
XX (IOWA ) UNIV IOWA RES FOUND.
XX
XX Krieg AM, Weiner G;
XX
XX WPI; 1999-620169/53.
DR
XX
PT Novel synergistic combinations of immunostimulatory oligonucleotides
PT and immunopotentiating cytokines are useful for stimulating the immune
PT system
XX
XX Example 8; Page 88; 91pp; English.
XX
XX Sequences AA241856-241949 are phosphorothioate CpG oligonucleotides
CC which are used in the invention to induce interleukin-12 (IL-12)
CC secretion from human PBMC. The invention comprises stimulating an immune
CC response in a subject comprising administering to a subject exposed to an
CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
CC oligonucleotide to induce a synergistic antigen specific immune
CC response. The methods are useful for treating cancer by stimulating an
CC antigen specific immune response against a cancer antigen. The methods
CC can also be used to treat neoplastic disorders in humans, including but
CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
CC for treating infectious diseases, e.g. viral diseases such as HIV,
CC bacterial diseases, and fungal diseases. The methods may also be used to
CC treat allergic diseases, e.g. asthma. The methods and compositions may
CC also be applied to treat cancer and tumours in non human subjects,
CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
CC caused by the bacterium Corynebacterium pseudotuberculosis, and
CC contagious lung tumour of sheep caused by jaagsiekte may also be
CC treated. CpG oligonucleotides can be useful in activating B cells, NK
CC cells, and antigen presenting cells, such as monocytes and macrophages.
CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
CC can be used as an adjuvant in conjunction with tumour antigens to
CC protect against a tumour challenge.
XX
SQ Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;

AA241945 Length: 20 March 5, 2002 14:20 Type: N Check: 5580
Found using 'seq2-3' (pappu403.key)

1 TCATACGCTTCCTAGCGTT
6 11 15 20
|-----| |-----|
|-----| |-----|

2 matches found in sequence:
aaz41946 ; IL-12 secretion inducing CpG oligonucleotide 91.
(from "mycobacterng.seq")
TOIG of: aaz41946 check: 5568 from: 1 to: 20

ID AA241946 standard; DNA; 20 BP.
XX
AC AA241946;
XX
XX 24-JAN-2000 (first entry)
XX
XX IL-12 secretion inducing CpG oligonucleotide 91.
XX
KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;

```

KW antigen presenting cell; infection; allergic disease.
 OS Synthetic.
 XX WO9951259-A2.
 PN 14-OCT-1999.
 PD 02-APR-1999; 99WO-US07335.
 XX 03-APR-1998; 98US-0080729.
 PR (IOWA) UNIV IOWA RES FOUND.
 XX Krieg AM, Weiner G;
 PI WPI; 1999-620169/53.
 DR Novel synergistic combinations of immunostimulatory oligonucleotides
 XX PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX Example 8; Page 88; 91pp; English.
 PS Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 XX which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;

AAZ41946 Length: 20 March 5, 2002 14:20 Type: N Check: 5568 ..
 Found using 'seq2-3' (pappu403.key)

```

1      |-----| |-----|
      6      11 15 20
      TCCATGACGTTCCCTGACGTT

```

 2 matches found in sequence:
 aaz41949 ; IL-12 secretion inducing CpG oligonucleotide 94.
 (from "mycobacterng.seq")
 TOIG of: aaz41949 check: 5568 from: 1 to: 20

ID AAZ41949 standard; DNA; 20 BP.
 XX AAZ41949;
 XX 24-JAN-2000 (first entry)
 DT IL-12 secretion inducing CpG oligonucleotide 94.
 XX

KW CpG oligonucleotide; phosphorothioate; interleukin-12; IL-12; secretion;
 KW human PBMC; immune response; cancer; HIV; bacterial disease; asthma;
 KW neoplastic disorder; jaagsiekte; B cell; NK cell; ss; cytokine;
 KW antigen presenting cell; infection; allergic disease.
 XX Synthetic.
 OS WO9951259-A2.
 PN 14-OCT-1999.
 PD 02-APR-1999; 99WO-US07335.
 XX 03-APR-1998; 98US-0080729.
 PR (IOWA) UNIV IOWA RES FOUND.
 XX Krieg AM, Weiner G;
 PI WPI; 1999-620169/53.
 DR Novel synergistic combinations of immunostimulatory oligonucleotides
 XX PT and immunopotentiating cytokines are useful for stimulating the immune
 PT system -
 XX Example 8; Page 89; 91pp; English.
 PS Sequences AAZ41856-241949 are phosphorothioate CpG oligonucleotides
 XX which are used in the invention to induce interleukin-12 (IL-12)
 CC secretion from human PBMC. The invention comprises stimulating an immune
 CC response in a subject comprising administering to a subject exposed to an
 CC antigen, an immunopotentiating cytokine and an immunostimulatory CpG
 CC oligonucleotide to induce a synergistic antigen specific immune
 CC response. The methods are useful for treating cancer by stimulating an
 CC antigen specific immune response against a cancer antigen. The methods
 CC can also be used to treat neoplastic disorders in humans, including but
 CC not limited to: sarcoma, carcinoma, fibroma, lymphoma, melanoma,
 CC neuroblastoma, retinoblastoma, and glioma. The methods are also useful
 CC for treating infectious diseases, e.g. viral diseases such as HIV,
 CC bacterial diseases, and fungal diseases. The methods may also be used to
 CC treat allergic diseases, e.g. asthma. The methods and compositions may
 CC also be applied to treat cancer and tumours in non human subjects,
 CC e.g. cats and dogs. Neoplasias affecting agricultural livestock may also
 CC be treated and include leukaemia, haemangiopericytoma and bovine ocular
 CC neoplasia. Chronic, infectious, contagious diseases of sheep and goats
 CC caused by the bacterium Corynebacterium pseudotuberculosis, and
 CC contagious lung tumour of sheep caused by jaagsiekte may also be
 CC treated. CpG oligonucleotides can be useful in activating B cells, NK
 CC cells, and antigen presenting cells, such as monocytes and macrophages.
 CC CpG oligonucleotides enhance antibody dependent cellular cytotoxicity and
 CC can be used as an adjuvant in conjunction with tumour antigens to
 CC protect against a tumour challenge.
 XX Sequence 20 BP; 3 A; 6 C; 4 G; 7 T; 0 other;

AAZ41949 Length: 20 March 5, 2002 14:20 Type: N Check: 5568 ..
 Found using 'seq2-3' (pappu403.key)

```

1      |-----| |-----|
      6      11 15 20
      TCCATGACGTTCCCTGACGTT

```

 6 matches found in sequence:
 aaz50938 ; Soluble Interleukin receptor-4 DNA (1).
 (from "mycobacterng.seq")
 TOIG of: aaz50938 check: 5767 from: 1 to: 2430

ID AAZ50938 standard; DNA; 2430 BP.
 XX AAZ50938;
 AC 05-JUN-2000 (first entry)
 XX

```

XX DE Soluble Interleukin receptor-4 DNA (I).
XX
XX Soluble Interleukin-4 receptor; sIL4-R; immune response;
KW immuno-deficiency; combination therapy; protease inhibitor;
KW immunosuppressive; immunoregulatory; infection; viral; bacterial; prion;
KW parasite; fungal; immunosuppression; neoplastic condition;
KW Human immunodeficiency virus; Mycobacterium; Salmonella; Pseudomonas;
KW Enterococcus; Vibrio; Plasmodium; Leishmania; Pneumocystis; Cryptococcus;
KW immunostimulator; antiviral; antibacterial; ss.
XX OS Mammalia.
XX
XX Key Location/Qualifiers
FH mat_peptide 1..2428
FT /*tag= a
FT /product= "Soluble IL-4 receptor"
FT /transl_except= (pos:226..228, aa:Ile)
XX
XX WO200009150-A2.
XX
XX 24-FEB-2000.
XX
XX 17-AUG-1999; 99WO-IB01432.
XX
XX 17-AUG-1998; 98US-0096735.
XX
XX (PREN/) PRENDERGAST P T.
XX
XX Prendergast PT;
XX
XX WPI; 2000-224174/19.
XX
XX P-PSDB; AAY70027.
XX
XX Novel methods of treating immunodeficiency resulting from viral and/or
XX bacterial infections comprising administration of Th2 cytokines with
XX Th2 antibodies, receptors, agonists or antagonists -
XX
XX Disclosure; Fig 1; 96pp; English.
XX
XX The present sequence encodes soluble Interleukin-4 receptor (sIL4-R)
XX used for enhancing immune response in conditions of immuno-deficiency by
XX a combination therapy comprising an IL-4 receptor and at least 1 agonist
XX and/or antagonist. Additionally the above combination may be
XX administered in combination with a protease inhibitor and/or reverse
XX transcriptase general anti-viral drugs. This therapy can provide an
XX immunosuppressive or immunoregulatory effect in a patient. The methods
XX are useful for treating immunodeficiency resulting from viral
XX bacterial, parasite, fungal, or prion infection, immunosuppression,
XX and/or a neoplastic condition in a patient, especially caused by Human
XX immunodeficiency virus, Mycobacterium, Salmonella, Pseudomonas,
XX Enterococcus, Vibrio, Plasmodium, Leishmania, Pneumocystis and
XX Cryptococcus.
XX
XX SQ Sequence 2430 BP; 528 A; 720 C; 675 G; 507 T; 0 other;
XX
XX AAZ50938 Length: 2430 March 5, 2002 14:18 Type: N Check: 5767
XX Found using 'seq2-3' (pappu403.key)
XX
XX
XX 98 CCGCTTCTCTGACTACATCCGACCTCCACGGTGTGAGTGGTTCTCGATAGCGCTGTGG
XX 148 153
XX
XX 158 ACTGCAGTTCTCAGCTCGCTACACTACACTACAGGCTGATGTTCTTCGA
XX
XX
XX 645 GTGGAGTCTTAGCATCAGGTGGTACAAACCACCTTCCAGCTGCCCTGATACAGCGCTTCC
XX 695 700
XX

```

```

705 ACTGGGGGTACCATCTCTGCTGCTCTGATCCCGTTGTTTGCCTG
...
1823 GCCTGCTCAGCAGCAATGGCATCCGCGGGGACACAGCAGCAGCGGGGACTGACGATGGC
1873 1878
1883 ATGGAGGCTACAAGCCTTCCAGAATCCTGTTCTTAACCAAGTCCCC
...
2150 GTGCCCACTGAAGCAACACCACAGCCAGGAGGAAGGTGGCCAGAGCCCATCCTTGTCTA
2200 2205
2210 GCCCTGGCTGTGGCTGCTGTACGATGACAGATCACCATCCCTGGG
...
2338 AACTTGTGAGGGAGGCAAGGGCCCTGGTCACTCTCTCTGTTCCAGCCAGACGACGAG
2388 2393
2398 GTGCCTGTGGCGCCCTGGGCATTGCTGTTTCT
2407 2412
-----
7 matches found in sequence:
aaz50939 ; Soluble interleukin receptor-4 DNA (II).
(from "mycobacterng.seq")
TOIG of: aaz50939 check: 4461 from: 1 to: 2520
ID AAZ50939 standard; DNA; 2520 BP.
XX AC AAZ50939;
XX DT 05-JUN-2000 (first entry)
XX DE Soluble Interleukin receptor-4 DNA (II).
XX KW Soluble Interleukin-4 receptor; sIL4-R; immune response;
XX immuno-deficiency; combination therapy; protease inhibitor;
XX immunosuppressive; immunoregulatory; infection; viral; bacterial; prion;
XX parasite; fungal; immunosuppression; neoplastic condition;
XX Human immunodeficiency virus; Mycobacterium; Salmonella; Pseudomonas;
XX Enterococcus; Vibrio; Plasmodium; Leishmania; Pneumocystis; Cryptococcus;
XX immunostimulator; antiviral; antibacterial; ss.
XX OS Mammalia.
XX FH Key Location/Qualifiers
FT mat_peptide 1..2520
FT /*tag= a
FT /product= "Soluble Interleukin-4 receptor"
FT /transl_except= (pos: 448..450, aa:Asp)
FT /transl_except= (pos: 1483..1485, aa: Leu)
FT /transl_except= (pos: 1945..1947, aa: Phe)
XX
XX PN WO200009150-A2.
XX
XX PD 24-FEB-2000.
XX
XX PF 17-AUG-1999; 99WO-IB01432.
XX
XX PR 17-AUG-1998; 98US-0096735.
XX
XX (PREN/) PRENDERGAST P T.
XX
XX Prendergast PT;
XX
XX WPI; 2000-224174/19.
XX
XX DR

```


CC treat sepsis. Extracellular IL-17RLP domains may be used as antagonists
CC of IL-17RLP. IL-17RLP agonists and antagonists can also be used to
CC modulate IL-6 expression, useful in treatment of cancers such as
CC myelomas, plasmacytomas and hybridomas and Lennert's Lymphoma. The
CC present cDNA sequence encodes human IL-17RLP. This cDNA sequence was
CC discovered in a cDNA library derived from human adult pulmonary tissue.
XX
SQ Sequence 1816 BP; 532 A; 439 C; 399 G; 445 T; 1 other;

AAZ52035 Length: 1816 March 5, 2002 14:18 Type: N Check: 7661
Found using 'seq2-3' (pappu403.key)

```
1 GCACGACGATGTCGCTGCTGCTAAAGCGTGGCCGCTGTGCAGGAGCGCGTACCC
  6 11 12 17
  |-----| |-----| |-----|
61 CGAGAGCCGACCGTTCATGTGCTCTGAACCTGGGCGCATCTCC
...
734 ATAGTGAAGGTGTACGGTGCAGCTGACTCCATATTTCTTCTACTTTGGCAGCAGTCCA
  |-----|
  784 789
794 TCCGACATAAAGAACAGTTGTGCTCTGCCACAAACAGCGTCCCTTCCCTCTGGATA
  |-----|
  832 837
854 ACAACAAAGCAAGCCGGAGGCTGGCTGCTCT
...
1133 AGATGGTCCAGTGCAGTGGCTTCCCACTCAAAGAGGAGCAGACAGCAAGTCTCTTCC
  |-----|
  1183 1188
1193 TTCTTTCCATAGCTCAACAGTGTGTGCGATGGTACCTGTGCAAGAGCGAGGCGAGTC
  |-----|
  1204 1209
1253 CCAGTGA
...
1320 TTCTCTGCACAAATACGTGGTGTCTACTTTAGAGAGATTGATACAAAGAGCATTACA
  |-----|
  1370 1375
1380 ATGCTCTCAGTGTCTGCCCAAGTACCACCTCATGAAGATGCCAC
...
1636 TTTTATACCAATAAAATTTTCAAAATATGCTAACTAATAGCATTAACGATTGGA
  |-----|
  1686 1691
1696 AACTACATTTTACAACCTTCAAAGCTGTTTATACATAGAAATCAATT
...
3 matches found in sequence:
aaz52037 : HETCC45RA cDNA clone related to IL-17RLP cDNA clone.
(from "mycobacterng.seq")
TOIG of: aaz52037 check: 8080 from: 1 to: 327
```

ID AAZ52037 standard; cDNA; 327 BP.
XX
AC AAZ52037;
XX
DT 09-AUG-2000 (first entry)

XX HETCC45RA cDNA clone related to IL-17RLP cDNA clone.
DE
XX
KW Interleukin-17-like receptor protein; IL-17RLP; cytokine receptor;
KW resistant chronic infection; acute infection; mycobacterial infection;
KW T-cell proliferation; IL-2 biosynthesis; lymphocytic leukaemia;
KW T-cell mediated autoimmune disease; hematopoiesis; sepsis; hybridoma;
KW IL-6 expression; myeloma; plasmacytoma; Lennert's Lymphoma;
KW Immunoprotective; cytostatic; hematopoietic; proliferative;
KW antibacterial; ss.
XX
OS Homo sapiens.
XX
XX WO2000015759-A1.
XX 23-MAR-2000.
XX
XX 15-SEP-1999; 99WO-US21048.
XX
XX 16-SEP-1998; 98US-0154219.
PR 16-SEP-1998; 98WO-US19121.
PR 16-MAR-1999; 99US-0268311.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Shi Y;
XX
XX WPI; 2000-271403/23.
XX
XX Novel polynucleotides encoding interleukin-17-like receptor protein,
XX useful for diagnosis and treatment of immune system-related disorders,
XX e.g. sepsis and cancers -
XX
XX Claim 23; Page 138; 147pp; English.
XX
XX The patent relates to novel interleukin-17-like receptor
XX protein (IL-17RLP). IL-17RLP is a homologue of the IL-17 receptor and has
XX a wide range of cytokine receptor-like activities. IL-17RLP or its
XX agonists may be used to enhance host defenses against resistant chronic
XX and acute infections, e.g. mycobacterial infections, via the attraction
XX and activation of microbicidal leukocytes. It may also be used to
XX increase T-cell proliferation by stimulating IL-2 biosynthesis, for the
XX treatment of T-cell mediated autoimmune diseases and lymphocytic
XX leukaemias. IL-17RLP may also be used to regulate hematopoiesis and to
XX treat sepsis. Extracellular IL-17RLP domains may be used as antagonists
XX of IL-17RLP. IL-17RLP agonists and antagonists can also be used to
XX modulate IL-6 expression, useful in treatment of cancers such as
XX myelomas, plasmacytomas and hybridomas and Lennert's Lymphoma. The
XX present sequence is the cDNA clone HETCC45RA related to human IL-17RLP
XX cDNA (ATCC deposit number 209198) which was discovered in a cDNA library
XX derived from human adult pulmonary tissue.
XX
XX Sequence 327 BP; 69 A; 63 C; 86 G; 76 T; 33 other;

AAZ52037 Length: 327 March 5, 2002 14:18 Type: N Check: 8080
Found using 'seq2-3' (pappu403.key)

```
1 AATTCGGCANAGCCCGCGATGTCGCTCGTGTGCTAGNCTNGNCGGCTGTNCAGGAGC
  16 21 22 27
  |-----| |-----| |-----|
61 GCCTGTACCCGAGAGCCGACCGTTCATGTGGCTCTGAACCTGGGNCATCTCC
  63
```

1 match found in sequence:
aaz52040 : 5'-primer for amplification of IL-17RLP leader sequence.
(from "mycobacterng.seq")

TOIG of: aaz52040 check: 3048 from: 1 to: 34

```
ID  AAZ52040 standard; DNA; 34 BP.
XX
AC  AAZ52040;
XX
DT  09-AUG-2000 (first entry)
XX
DE  5' primer for amplification of IL-17RLP leader sequence.
XX
KW  Interleukin-17-like receptor protein; IL-17RLP; cytokine receptor;
KW  resistant chronic infection; acute infection; mycobacterial infection;
KW  T-cell proliferation; IL-2 biosynthesis; lymphocytic leukaemia;
KW  T-cell mediated autoimmune disease; hematopoiesis; sepsis; hybridoma;
KW  IL-6 expression; myeloma; plasmacytoma; Lennert's lymphoma;
KW  immunoprotective; cytostatic; hematopoietic; proliferative;
KW  antibacterial; PCR primer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200015759-A1.
XX
PD  23-MAR-2000.
XX
PF  15-SEP-1999; 99WO-US21048.
XX
PR  16-SEP-1998; 98US-0154219.
PR  16-SEP-1998; 98WO-US19121.
PR  16-MAR-1999; 99US-0268311.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Ruben SM, Shi Y;
XX
WPI: 2000-271403/23.
XX
PT  Novel polynucleotides encoding interleukin-17-like receptor protein,
PT  useful for diagnosis and treatment of immune system-related disorders,
PT  e.g. sepsis and cancers -
XX
PS  Example 2; Page 98; 147pp; English.
XX
CC  The patent relates to novel interleukin-17-like receptor
CC  protein (IL-17RLP). IL-17RLP is a homologue of the IL-17 receptor and has
CC  a wide range of cytokine receptor-like activities. IL-17RLP or its
CC  agonists may be used to enhance host defenses against resistant chronic
CC  and acute infections, e.g. mycobacterial infections, via the attraction
CC  and activation of microbicidal leukocytes. It may also be used to
CC  increase T-cell proliferation by stimulating IL-2 biosynthesis, for the
CC  treatment of T-cell mediated autoimmune diseases and lymphocytic
CC  leukaemias. IL-17RLP may also be used to regulate hematopoiesis and to
CC  treat sepsis. Extracellular IL-17RLP domains may be used as antagonists
CC  of IL-17RLP. IL-17RLP agonists and antagonists can also be used to
CC  modulate IL-6 expression, useful in treatment of cancers such as
CC  myelomas, plasmacytomas and hybridomas and Lennert's lymphoma. The
CC  present sequence is the 5' PCR primer used for the amplification of
CC  IL-17RLP leader sequence. This is used in the cloning and expression of
CC  IL-17RLP protein in a baculovirus expression system.
XX
SQ  Sequence 34 BP; 4 A; 11 C; 11 G; 8 T; 0 other;
```

AAZ52040 Length: 34 March 5, 2002 14:18 Type: N Check: 3048 ..
Found using 'seq2-3' (pappu403.key)

```
1  CGCGGATCCATGTCGCTCGTGTGCTGAAGCCTGG
    12 17
    |-----|
    GCGGGATCCGCCACCACCTCTTCTCCACAGGCGCTTTCGGTCCAGTTGCCTTCTCC
```

2 matches found in sequence:
aaz52042; 5' primer for amplification of IL-17RLP cDNA.
(from "mycobacterng.seq")
TOIG of: aaz52042 check: 8430 from: 1 to: 129

```
ID  AAZ52042 standard; DNA; 129 BP.
XX
AC  AAZ52042;
XX
DT  09-AUG-2000 (first entry)
XX
DE  5' primer for amplification of IL-17RLP cDNA.
XX
KW  Interleukin-17-like receptor protein; IL-17RLP; cytokine receptor;
KW  resistant chronic infection; acute infection; mycobacterial infection;
KW  T-cell proliferation; IL-2 biosynthesis; lymphocytic leukaemia;
KW  T-cell mediated autoimmune disease; hematopoiesis; sepsis; hybridoma;
KW  IL-6 expression; myeloma; plasmacytoma; Lennert's lymphoma;
KW  immunoprotective; cytostatic; hematopoietic; proliferative;
KW  antibacterial; PCR primer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200015759-A1.
XX
PD  23-MAR-2000.
XX
PF  15-SEP-1999; 99WO-US21048.
XX
PR  16-SEP-1998; 98US-0154219.
PR  16-SEP-1998; 98WO-US19121.
PR  16-MAR-1999; 99US-0268311.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Ruben SM, Shi Y;
XX
WPI: 2000-271403/23.
XX
PT  Novel polynucleotides encoding interleukin-17-like receptor protein,
PT  useful for diagnosis and treatment of immune system-related disorders,
PT  e.g. sepsis and cancers -
XX
PS  Example 3; Page 102; 147pp; English.
XX
CC  The patent relates to novel interleukin-17-like receptor
CC  protein (IL-17RLP). IL-17RLP is a homologue of the IL-17 receptor and has
CC  a wide range of cytokine receptor-like activities. IL-17RLP or its
CC  agonists may be used to enhance host defenses against resistant chronic
CC  and acute infections, e.g. mycobacterial infections, via the attraction
CC  and activation of microbicidal leukocytes. It may also be used to
CC  increase T-cell proliferation by stimulating IL-2 biosynthesis, for the
CC  treatment of T-cell mediated autoimmune diseases and lymphocytic
CC  leukaemias. IL-17RLP may also be used to regulate hematopoiesis and to
CC  treat sepsis. Extracellular IL-17RLP domains may be used as antagonists
CC  of IL-17RLP. IL-17RLP agonists and antagonists can also be used to
CC  modulate IL-6 expression, useful in treatment of cancers such as
CC  myelomas, plasmacytomas and hybridomas and Lennert's lymphoma. The
CC  present sequence is the 5' PCR primer used for the amplification of
CC  IL-17RLP cDNA. This primer comprises a BamHI site, a kozak sequence, a
CC  start codon, and 25 nucleotides of the 5' coding region of the
CC  extracellular domain of the IL-17RLP protein. This is used in the
CC  construction of vectors for expression in E. coli.
XX
SQ  Sequence 129 BP; 13 A; 50 C; 32 G; 34 T; 0 other;
```

AAZ52042 Length: 129 March 5, 2002 14:18 Type: N Check: 8430 ..
Found using 'seq2-3' (pappu403.key)

```
1  GCGGGATCCGCCACCACCTCTTCTCCACAGGCGCTTTCGGTCCAGTTGCCTTCTCC
    34 39
    |-----|
    CTGGGGCTGCTCCTCTGTGTGCTGTGCTGTCCCTGCCGCCAGTATGTCCTCGTGTGC
```

61 CTGGGGCTGCTCCTCTGTGTGCTGTGCTGTCCCTGCCGCCAGTATGTCCTCGTGTGC
107 112
|-----|

121 TAAGCCTGG

1 match found in sequence:

aa52044 ; 5' primer for cloning IL-17RLP extracellular domain in CHO cells.
(from "mycobacterng.seq")
TOIG of: aa52044 check: 7995 from: 1 to: 43

ID AAZ52044 standard; DNA; 43 BP.

XX AC AAZ52044;

XX XX 09-AUG-2000 (first entry)

XX DE 5' primer for cloning IL-17RLP extracellular domain in CHO cells.

XX KW Interleukin-17-like receptor protein; IL-17RLP; cytokine receptor;
KW resistant chronic infection; acute infection; mycobacterial infection;
KW T-cell proliferation; IL-2 biosynthesis; lymphocytic leukaemia;
KW T-cell mediated autoimmune disease; hematopoiesis; sepsis; hybridoma;
KW IL-6 expression; myeloma; plasmacytoma; Lennert's Lymphoma;
KW Immunoprotective; cytostatic; hematopoietic; proliferative;
KW antibacterial; PCR primer; CHO; chinese hamster ovary; ss.

XX OS Homo sapiens.

XX PN WO200015759-A1.

XX XX 23-MAR-2000.

XX PF 15-SEP-1999; 99WO-US21048.

XX PR 16-SEP-1998; 98US-0154219.

XX PR 16-SEP-1998; 98WO-US19121.

XX PR 16-MAR-1999; 99US-0268311.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX XX Ruben SM, Shi Y;

XX XX WPI; 2000-271403/23.

XX PT Novel polynucleotides encoding interleukin-17-like receptor protein,
XX useful for diagnosis and treatment of immune system-related disorders,
XX e.g. sepsis and cancers -

XX PS Example 3; Page 104; 147pp; English.

XX CC The patent relates to novel interleukin-17-like receptor
XX protein (IL-17RLP). IL-17RLP is a homologue of the IL-17 receptor and has
XX a wide range of cytokine receptor-like activities. IL-17RLP or its
XX agonists may be used to enhance host defenses against resistant chronic
XX and acute infections, e.g. mycobacterial infections, via the attraction
XX and activation of microbicidal leukocytes. It may also be used to
XX increase T-cell proliferation by stimulating IL-2 biosynthesis, for the
XX treatment of T-cell mediated autoimmune diseases and lymphocytic
XX leukaemias. IL-17RLP may also be used to regulate hematopoiesis and to
XX treat sepsis. Extracellular IL-17RLP domains may be used as antagonists
XX of IL-17RLP. IL-17RLP agonists and antagonists can also be used to
XX modulate IL-6 expression, useful in treatment of cancers such as
XX myelomas, plasmacytomas and hybridomas and Lennert's Lymphoma. The
XX present sequence is the 5' PCR primer used for the cloning of the
XX extracellular domain of IL-17RLP and expression in CHO (chinese hamster
XX ovary) cells. This primer comprises a BamHI site, Kozak sequence, a start
XX codon, and 25 nucleotides of the 5' coding region of the extracellular
XX domain of the IL-17RLP protein.

XX SQ Sequence 43 BP; 6 A; 16 C; 12 G; 9 T; 0 other;

AAZ52044 Length: 43 March 5, 2002 14:18 Type: N Check: 7995
Found using 'seq2-3' (pappu403.key)

|-----|

1 CTAGCCGGATCGCCACCATGTCGCTGCTGCTAAGCCTGG
21 26

9 matches found in sequence:

aa52046 ; CDNA encoding interleukin 17-like receptor protein (IL17RLP)-2.
(from "mycobacterng.seq")
TOIG of: aa52046 check: 135 from: 1 to: 1918

ID AAZ52046 standard; cDNA; 1918 BP.

XX XX

XX AC AAZ52046;

XX XX 09-AUG-2000 (first entry)

XX DE cDNA encoding interleukin 17-like receptor protein (IL17RLP)-2.

XX KW Interleukin-17-like receptor protein; IL-17RLP; cytokine receptor;
KW resistant chronic infection; acute infection; mycobacterial infection;
KW T-cell proliferation; IL-2 biosynthesis; lymphocytic leukaemia;
KW T-cell mediated autoimmune disease; hematopoiesis; sepsis; hybridoma;
KW IL-6 expression; myeloma; plasmacytoma; Lennert's Lymphoma;
KW Immunoprotective; cytostatic; hematopoietic; proliferative;
KW antibacterial; ss.

XX OS Homo sapiens.

XX XX Key Location/Qualifiers

XX FH IL1..1412

XX FT CDS

XX FT /tag= a /product= "Interleukin-17 like receptor protein"

XX XX WO200015759-A1.

XX XX 23-MAR-2000.

XX XX 15-SEP-1999; 99WO-US21048.

XX XX 16-SEP-1998; 98US-0154219.

XX PR 16-SEP-1998; 98WO-US19121.

XX PR 16-MAR-1999; 99US-0268311.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX XX Ruben SM, Shi Y;

XX XX WPI; 2000-271403/23.

XX XX P-PSDB; AAY70595.

XX PT Novel polynucleotides encoding interleukin-17-like receptor protein,
XX useful for diagnosis and treatment of immune system-related disorders,
XX e.g. sepsis and cancers -

XX PS Disclosure; Page 141-143; 147pp; English.

XX CC The patent relates to novel interleukin-17-like receptor
XX protein (IL-17RLP). IL-17RLP is a homologue of the IL-17 receptor and has
XX a wide range of cytokine receptor-like activities. IL-17RLP or its
XX agonists may be used to enhance host defenses against resistant chronic
XX and acute infections, e.g. mycobacterial infections, via the attraction
XX and activation of microbicidal leukocytes. It may also be used to
XX increase T-cell proliferation by stimulating IL-2 biosynthesis, for the
XX treatment of T-cell mediated autoimmune diseases and lymphocytic
XX leukaemias. IL-17RLP may also be used to regulate hematopoiesis and to
XX treat sepsis. Extracellular IL-17RLP domains may be used as antagonists
XX of IL-17RLP. IL-17RLP agonists and antagonists can also be used to
XX modulate IL-6 expression, useful in treatment of cancers such as
XX myelomas, plasmacytomas and hybridomas and Lennert's Lymphoma. The
XX present cDNA sequence encodes human IL-17RLP. This cDNA sequence was
XX derived from sequencing the HAPOR40 cDNA clone (ATCC deposit
XX number: 209198)

XX SQ Sequence 1918 BP; 560 A; 469 C; 426 G; 461 T; 2 other;

AAZ52046 Length: 1918 March 5, 2002 14:18 Type: N Check: 135
Found using 'seq2-3' (pappu403.key)

```
...
57  GCGGCGGCTCTAGAACTAGTGGATCCCGGGCTGCAGGAATTCNGCAGCGATGTCG      |----|
    107 112 113
-|
117  CTCGTGCTGCTAAGCCTGGCGGCTGTGCAGGAGCGCGTACCCGAGAGCGGACCGTT      |----|
    118 150 155
177  CAATGTGGCTCTGAAACTGGGCCATCTCC
...
835  ATAGTGAAGGTGTACGGTGCAGCTGACGCCATATTTTCCTACTTGTGGCAGCGACTGCA      |----|
    885 890
895  TCCGACATAAGGAACAGTGTGCTCTGCCACAAACAGCGTCCTTTCCCTCTGGATA      |----|
    933 938
955  ACAACAAAAGCAAGCGGGAGGCTGGCTGCCTCT
...
1234 AGATGGGTCCAGTGCAGTGGCTTGCCACTCAAAGAAGCGCAGACAAAGTCGTCTTCC      |----|
    1284 1289
1294 TTCTTTCCAATGACGTCAACAGTGTGTGGGATGGTACCTGTGCAAGAGCGAGGGCAGTC      |----|
    1305 1310
1354 CCAGTGA
...
1422 TTCATCTGCACAAATACGTGGTGGTCTACTTTAGAGAGATTGATACAAAAGACGATTACA      |----|
    1472 1477
1482 ATGCTCTCAGTGTCTGCCCAAGTACCACCTCATGAAGGATGCCAC
...
1738 TTTTATACCAATAAAATTTTCAAATATGCTAACTAATGTAGCAATTAACGATTGGA      |----|
    1788 1793
1798 AACTACATTTACAACCTCAAAGCTGTTTATACATAGAAATCAATT
...
```

-- Search Statistics --

| | | |
|--------------------------------|-------------|---------------|
| Times: | CPU | Total Elapsed |
| | 00:00:04.16 | 00:00:17.00 |
| Number of sequences searched: | | 880 |
| Number of sequence hits: | | 482 |
| Number of separate matches: | | 3982 |
| Number of sequence hits saved: | | 0 |


```
> O <
O| |O Intelligenetics
> O <

Quest - Quick User-directed Expression Search Tool
Release 5.4

-- Outline of search "seq4est" --

Selected search type is key against sequence data banks or files.
Selected scope is Sequence.
Selected sequence key from "pappu403.key":
seq4 (NA) ID seq4 NA preliminary pattern
1 either-or
2 aacgttcg
2 cgaacgtt

Selected files:
File : immun_est.seq

-- Output Parameters --

Format Options:
Nucleic acid code matching Exact Indirect file No
Find non-matching hits only No Sequence or key file No
Report key used Yes List of hits Yes
Note position of hit Yes Hit display Yes
Display full annotations Yes Name and annotations Yes
Sequence context 50

-- Run Parameters --

Run mode Batch
Time to start comparison now
Notify at end of run No

No hits found.

-- Search Statistics --

Times:
CPU 00:00:00.01 Total Elapsed 00:00:01.00

Number of sequences searched: 119
Number of sequence hits: 0
Number of separate matches: 0
Number of sequence hits saved: 0
```


> O <
OI IO Intelligenetics
> O <

Quest - Quick User-directed Expression Search Tool
Release 5.4

-- Outline of search "Variant2-3est" --

Selected search type is key against sequence data banks or files.

Selected scope is Sequence.

Selected sequence key from "pappu403.key":

seq2-3 (NA) ID seq2-3 NA preliminary pattern

either-or

followed by

a or g

a or g

cg

c or t

c or t

c or t

followed by

a or g

a or g

cga

c or t

Selected files:

File : immun_est.seq

-- Output Parameters --

Format Options:

| Nucleic acid code matching | Exact | Indirect file |
|-----------------------------|-------|---------------|
| Find non-matching hits only | No | No |
| Report key used | Yes | Yes |
| Note position of hit | Yes | Yes |
| Display full annotations | Yes | Yes |
| Sequence context | 50 | |

-- Run Parameters --

| Run mode | Batch |
|--------------------------|-------|
| Time to start comparison | now |
| Notify at end of run | No |

5 matches found in sequence:
au090800 ; TOIG of: au090800 check: 4823 from: 1 to: 1040
(from "immun_est.seq")
TOIG of: au090800 check: 4823 from: 1 to: 1040

| | | | | | |
|------------|--|-------------|------|-----|-------------|
| LOCUS | AU090800 | 1040 bp | mRNA | EST | 23-JAN-2001 |
| DEFINITION | AU090800 lambda ZAPII-Con A stimulated leukocytes Paralicthys olivaceus cDNA clone JFCOnA212R reverse similar to Homo sapiens T-cell immune regulator (AF025374), mRNA sequence. | | | | |
| ACCESSION | AU090800 | | | | |
| VERSION | AU090800.1 | GI:12390840 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | bustard halibut. | | | | |
| ORGANISM | Paralicthys olivaceus | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes; Pleuronectoidae; Bothidae; Paralicthys. | | | | |
| REFERENCE | 1 (bases 1 to 1040) | | | | |

AUTHORS Nam,B., Hirono,I. and Aoki,T.
TITLE Identification of expressed genes from Japanese flounder (Paralicthys olivaceus) leukocytes stimulated with Con A/PMA Unpublished (2000)
JOURNAL
COMMENT Contact: Ikuo Hirono
Laboratory of Genetics and Biochemistry
Tokyo University of Fisheries
Konan 4-5-7, Minato-ku, Tokyo 108, Japan
Email: hirono@tokyo-u-fish.ac.jp.
FEATURES
source Location/Qualifiers
1..1040
/organism="Paralicthys olivaceus"
/db_xref="taxon:8255"
/clone_lib="JFCOnA212R"
/clone_lib="lambda ZAPII-Con A stimulated leukocytes"
/cell_type="leukocytes"
/dev_stage="adult"
BASE COUNT 328 a 214 c 248 g 250 t
ORIGIN
AU090800 Length: 1040 February 27, 2002 16:11 Type: N Check: 4823 ..
Found using 'seq2-3' (pappu403.key)

...
153 GTGTTCAAGACATAGCTACTGATTCTGAAATGACAAGAGGTTCAAAACAGCGACAGTC
203 208
213 TGACAGATAATCTTGGCTCAGTCTCAGAGAAAGTGTGTAATAATGCTTC
...
280 ACCACACAAATTCAGATGAGAGATAGTAATTTGAAGTAGAAGAAGGCGTCAGAT
330 335
340 CTCAGTCTTCAAACCTCCCGTCCAAGATGCTTTTCGAAGGTGAAGGG
...
498 GTGAGAACAGCAAGAAGTAAAGACTATGGCCAGAGGAAGAAGCTCAACGCTGCTG
548 553
558 GAAGAAAGGCCGATGTGCATCACCATGGTCCACAGCACCTCCGACA
...
858 CTCGCGAGATCAAGTCTTTCACATATTACATGCAGGGAACACATGCCAAGGCGATAACC
908 913
918 ACTAAGAGTGATTGTACACCGCTCTCCACGGTAGAGAGGTTTGTACTGGGTCGTTG
971 976
978 TTGTTGAAGAGAACATGTTGATAAAGGCCAATGAGGAGACTGGGAGCTC
...

1 match found in sequence:
h39313 ; TOIG of: h39313 check: 1615 from: 1 to: 170
(from "immun_est.seq")
TOIG of: h39313 check: 1615 from: 1 to: 170
LOCUS H39313 170 bp mRNA EST
DEFINITION DR3 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone DR3. similar to tyrosine phosphatase (PTP-BAS, type 3), mRNA sequence.
ACCESSION H39313
VERSION H39313.1 GI:915103

```

KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 170)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
JOURNAL rat astrocytes by PCR differential display
MEDLINE Neuroimmunomodulation 2, 347-355 (1995)
COMMENT 96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
high similarity to human tyrosine phosphatase (PTP-BAS, type 3)
mRNA (D21211)
subcloned into pGEM-T
Seq primer: RP.

FEATURES
source
Location/Qualifiers
1..170
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR3"
/clone.lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."
BASE COUNT 46 a 40 c 33 g 51 t
ORIGIN
H39313 Length: 170 February 27, 2002 16:11 Type: N Check: 1615
Found using 'seq2-3' (pappu403.key)
1 CTACCTTTCTTGTGCTCTCTTCGCGTTCGCGACAACTAAATGAACAAAAGG
34 39
61 CTTAATTAAAGTACTACTACATCA
...

-----
1 match found in sequence:
h39314 ; TOIG of: h39314 check: 7333 from: 1 to: 166
(from "immun_est.seq")
TOIG of: h39314 check: 7333 from: 1 to: 166

LOCUS H39314 166 bp mRNA EST 27-AUG-1996
DEFINITION DR4 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
H39314
ACCESSION H39314
VERSION H39314.1 GI:915104
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 166)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
JOURNAL rat astrocytes by PCR differential display
MEDLINE Neuroimmunomodulation 2, 347-355 (1995)
COMMENT 96437728
Contact: Hart RP
Rutgers University

KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 170)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
JOURNAL rat astrocytes by PCR differential display
MEDLINE Neuroimmunomodulation 2, 347-355 (1995)
COMMENT 96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
high similarity to human tyrosine phosphatase (PTP-BAS, type 3)
mRNA (D21211)
subcloned into pGEM-T
Seq primer: RP.

FEATURES
source
Location/Qualifiers
1..170
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR3"
/clone.lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."
BASE COUNT 46 a 40 c 33 g 51 t
ORIGIN
H39313 Length: 170 February 27, 2002 16:11 Type: N Check: 1615
Found using 'seq2-3' (pappu403.key)
1 CTACCTTTCTTGTGCTCTCTTCGCGTTCGCGACAACTAAATGAACAAAAGG
34 39
61 CTTAATTAAAGTACTACTACATCA
...

-----
1 match found in sequence:
h39314 ; TOIG of: h39314 check: 7333 from: 1 to: 166
(from "immun_est.seq")
TOIG of: h39314 check: 7333 from: 1 to: 166

LOCUS H39314 166 bp mRNA EST 27-AUG-1996
DEFINITION DR4 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
H39314
ACCESSION H39314
VERSION H39314.1 GI:915104
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 166)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
JOURNAL rat astrocytes by PCR differential display
MEDLINE Neuroimmunomodulation 2, 347-355 (1995)
COMMENT 96437728
Contact: Hart RP
Rutgers University

```

```

Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
sequence is highly similar to rat plasma membrane
Ca2+ATPase-isoform 1 mRNA (J03753)
subcloned into PCRscript
Seq primer: AURE.

FEATURES
source
Location/Qualifiers
1..166
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR4"
/clone.lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."
BASE COUNT 46 a 39 c 38 g 43 t
ORIGIN
H39314 Length: 166 February 27, 2002 16:11 Type: N Check: 7333
Found using 'seq2-3' (pappu403.key)
1 GCATTACGTAGACACCTGGACTCGAGGTACTCCCGGCGTCGAGCGGTTACACTACGGT
36 41
61 TTCACGGGTAGTACTAATCGAAGAAAGTTTC
...

-----
1 match found in sequence:
h39320 ; TOIG of: h39320 check: 6165 from: 1 to: 171
(from "immun_est.seq")
TOIG of: h39320 check: 6165 from: 1 to: 171

LOCUS H39320 171 bp mRNA EST 27-AUG-1996
DEFINITION DR16 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
H39320
ACCESSION H39320
VERSION H39320.1 GI:915110
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 171)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
JOURNAL rat astrocytes by PCR differential display
MEDLINE Neuroimmunomodulation 2, 347-355 (1995)
COMMENT 96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
sequence is highly similar to rat mRNA for ATP synthase subunit e
(D13121)
subcloned into PCRscript
Seq primer: KS.

FEATURES
source
Location/Qualifiers
1..171
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

```

/clone="DR16"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."

BASE COUNT 56 a 20 c 56 g 39 t

H39320 Length: 171 February 27, 2002 16:11 Type: N Check: 6165
Found using 'seq2-3' (pappu403.key)

...
87 GAAGATGTCAGCATTTTCAAGTGGCGGCGAGTCATCTTCTTTCTCTTCTGCTTGAGG
137 142

147 ATGAATAAAGCTTCATTGTCTTGAA

1 match found in sequence:
h39321 : TOIG of: h39321 check: 9742 from: 1 to: 86
(from "immun_est.seq")
TOIG of: h39321 check: 9742 from: 1 to: 86

LOCUS H39321 86 bp mRNA EST 27-AUG-1996
DEFINITION DR20 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR20 similar to complement C1 inhibitor, mRNA sequence.

ACCESSION H39321
VERSION H39321.1 GI:915111
KEYWORDS EST.
SOURCE Norway rat.

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 86)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE 9643728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102

FEATURES
source 1..86
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR20"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."
Seq primer: KS.

BASE COUNT 17 a 25 c 25 g 19 t

H39321 Length: 86 February 27, 2002 16:11 Type: N Check: 9742
Found using 'seq2-3' (pappu403.key)

...
1 GAGATGAGTGGCTCTGTTGCCAGGCTGGAGTGCAATGATGCAATCTCGGCTCACTGCA

2 matches found in sequence:
h39327 : TOIG of: h39327 check: 6825 from: 1 to: 189
(from "immun_est.seq")

9 14

61 AGCT

...

2 matches found in sequence:
h39324 : TOIG of: h39324 check: 2559 from: 1 to: 163
(from "immun_est.seq")
TOIG of: h39324 check: 2559 from: 1 to: 163

LOCUS H39324 163 bp mRNA EST 27-AUG-1996
DEFINITION DR33 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR33 similar to unknown sequence, mRNA sequence.

ACCESSION H39324
VERSION H39324.1 GI:915114
KEYWORDS EST.
SOURCE Norway rat.

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 163)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE 9643728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102

FEATURES
source 1..163
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR33"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."
Seq primer: KS.

BASE COUNT 44 a 41 c 39 g 39 t

H39324 Length: 163 February 27, 2002 16:11 Type: N Check: 2559
Found using 'seq2-3' (pappu403.key)

1 GGTAAATGGCTGCTAGTCTGCTACTTCTTACGCACCTTAAGCAAAGTGGGCGATGAAGA
49 54

61 AGCAATCAGCTAGTATTAGCAGCGCTCCAGCGCTCTACTGCATATTACGAGGGCTTATCT
79 84

121 CAAAGCCAATCTCT

...

2 matches found in sequence:
h39327 : TOIG of: h39327 check: 6825 from: 1 to: 189
(from "immun_est.seq")

TOIG of: h39327 check: 6825 from: 1 to: 189

LOCUS H39327 189 bp mRNA EST 27-AUG-1996
 DEFINITION DR46 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
 DR46 similar to LINE repetitive element, mRNA sequence.
 ACCESSION H39327
 VERSION H39327.1 GI:915117
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 189)
 AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
 TITLE Identification of mRNAs regulated by interferon-gamma in cultured
 rat astrocytes by PCR differential display

JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
 MEDLINE 96437728
 COMMENT Contact: Hart RP
 Rutgers University
 Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
 07102
 Tel: 2016481008
 Fax: 2016481007
 Email: rhart@andromeda.rutgers.edu

sequence contains rat repetitive sequence similar to 3'-end of LINE
 repetitive element (M3922).
 subcloned into PCRscript

Seq primer: KS.

FEATURES
 source Location/Qualifiers
 1..189

/organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="DR46"

/clone_lib="IFNgamma-induced astrocyte ESTs"
 /note="cDNA from IFNgamma-induced cultured rat astrocytes
 was screened by PCR differential display. Regulated cDNAs
 were amplified, sequenced and confirmed by dot-blot and
 Northern." Northern.

BASE COUNT 29 a 45 c 50 g 65 t
 ORIGIN

H39327 Length: 189 February 27, 2002 16:11 Type: N Check: 6825
 Found using 'seq2-3' (pappu403.key)

...

66 ATTAGAGTCGATGAGTCCCGACTCGGTCTCTTCTAGTAATTGGTGGTCGTCGTCCTCCA
 116 121 125
 |-----|

126 ACGTCACTCGACTCTAGTATACGGTGACGTGAGATCGGACCCGCTGTCTCACTCT
 130

...

1 match found in sequence:
 h39330 ; TOIG of: h39330 check: 9318 from: 1 to: 189
 (from "immun_est.seq")
 TOIG of: h39330 check: 9318 from: 1 to: 189

LOCUS H39330 189 bp mRNA EST 27-AUG-1996
 DEFINITION DR76 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
 DR76 similar to proto-oncogene tyrosine phosphatase, mRNA sequence.
 ACCESSION H39330
 VERSION H39330.1 GI:915120
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 AUTHORS

TITLE 1 (bases 1 to 189)
 Kuchinke,W., Hart,R.P. and Jonakait,G.M.
 Identification of mRNAs regulated by interferon-gamma in cultured
 rat astrocytes by PCR differential display

JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
 MEDLINE 96437728
 COMMENT Contact: Hart RP
 Rutgers University
 Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
 07102
 Tel: 2016481008
 Fax: 2016481007
 Email: rhart@andromeda.rutgers.edu

highest similarity to human proto-oncogene tyrosine phosphatase
 (U07563)
 subcloned into PCRscript

Seq primer: KS.

FEATURES
 source Location/Qualifiers
 1..189

/organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="DR76"

/clone_lib="IFNgamma-induced astrocyte ESTs"
 /note="cDNA from IFNgamma-induced cultured rat astrocytes
 was screened by PCR differential display. Regulated cDNAs
 were amplified, sequenced and confirmed by dot-blot and
 Northern." Northern.

BASE COUNT 29 a 45 c 50 g 65 t
 ORIGIN

H39330 Length: 189 February 27, 2002 16:12 Type: N Check: 9318
 Found using 'seq2-3' (pappu403.key)

1 TTGAGACAGTCTCACTGTGCGCCAGGCTAGAGTCAGTGGCATATGATCTCAGTCCAC
 19 24
 |-----|

61 TGCAACCTCTGCTG

...

3 matches found in sequence:
 h39336 ; TOIG of: h39336 check: 927 from: 1 to: 65
 (from "immun_est.seq")
 TOIG of: h39336 check: 927 from: 1 to: 65

LOCUS H39336 65 bp mRNA EST 27-AUG-1996
 DEFINITION DR14 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
 DR14 5'-end with RP similar to 28S ribosomal RNA, mRNA sequence.

ACCESSION H39336
 VERSION H39336.1 GI:915010
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 65)

AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
 TITLE Identification of mRNAs regulated by interferon-gamma in cultured
 rat astrocytes by PCR differential display

JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
 MEDLINE 96437728
 COMMENT Contact: Hart RP
 Rutgers University
 Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
 07102
 Tel: 2016481008

Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
sequence is highly similar to rat 28S ribosomal RNA (K01591).
direct sequence
Seq primer: CCGAGTTCC.
Location/Qualifiers
1. .65
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR14"
/note="cDNA from IFNgamma-induced astrocyte ESTs"
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern." 11 a 22 c 17 g 15 t

BASE COUNT 11 a 22 c 17 g 15 t
ORIGIN

H39336 Length: 65 February 27, 2002 16:12 Type: N Check: 927
Found using 'seq2-3' (pappu403.key)

1 ATCGGCGCTTAACCGCGTTCGTTTCATCCGCGCCAGTTCGTTACCAAAAG
5 10 18 23 37 42
61 TGGCT

3 matches found in sequence:
h39341 ; TOIG of: h39341 check: 3470 from: 1 to: 157
(from "immun_est.seq")
TOIG of: h39341 check: 3470 from: 1 to: 157

LOCUS H39341 157 bp mRNA EST 27-AUG-1996
DEFINITION DR24 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR24 3'-end with RP similar to ribosomal RNA, mRNA sequence.
ACCESSION H39341
VERSION H39341.1 GI:915015
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 157)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE 9643728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
sequence is highly similar to rat ribosomal RNA (V01270)
subcloned into PCR script
Seq primer: RP.
Location/Qualifiers
1. .157
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR24"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes"
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern." 39 a 43 c 40 g 35 t

BASE COUNT 39 a 43 c 40 g 35 t

ORIGIN

H39341 Length: 157 February 27, 2002 16:12 Type: N Check: 3470
Found using 'seq2-3' (pappu403.key)

1 GGATAGCTGGCGCTCTCGCAACGCGTCGCTCGACACCGCGAGTTTATCCGCGTAAAGC
9 14 26 31
61 GAATGATTAGAGCTCTTGGGCGCGCAACGATCTCAACCTATTCTCAAACTTTAAATGGGT
86 91
121 AAGAAGCCGCGCTTGGGCTGC
...

2 matches found in sequence:
h39349 ; TOIG of: h39349 check: 76 from: 1 to: 176
(from "immun_est.seq")
TOIG of: h39349 check: 76 from: 1 to: 176

LOCUS H39349 176 bp mRNA EST 27-AUG-1996
DEFINITION DR32 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR32 5'-end with RP similar to unknown sequence, mRNA sequence.
ACCESSION H39349
VERSION H39349.1 GI:915023
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 176)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE 9643728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into PCRscript
Seq primer: KS.
Location/Qualifiers
1. .176
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR32"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes"
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern." 45 a 43 c 44 g 44 t

BASE COUNT 45 a 43 c 44 g 44 t
ORIGIN

H39349 Length: 176 February 27, 2002 16:12 Type: N Check: 76
Found using 'seq2-3' (pappu403.key)

...

58 AGCAACTTTCACACACAAACGAAATGGCTGCCCAACAGCAGGAGGAGAAATCGTCTGTGG
108 113

2

07102

Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into PCRscript
Seq primer: KS.

FEATURES

source

Location/Qualifiers

1. .155
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR45"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern." 49 a 43 c 30 g 33 t

BASE COUNT

ORIGIN

H39359 Length: 155 February 27, 2002 16:12 Type: N Check: 4609 ..
Found using 'seq2-3' (pappu403.key)

...

18 CCTGCCTACTCTTACCAGGAATGTGGTCCAGGATAGTGACACCAACGTCAGAC
68 73
|-----|

78 AACAGTGGTGGGAACAAATCACACATATATACAGCGGGACACTA

...

2 matches found in sequence:

h39361 ; TOIG of: h39361 check: 3789 from: 1 to: 166
(from "immun_est.seq")
TOIG of: h39361 check: 3789 from: 1 to: 166

LOCUS H39361 166 bp mRNA EST 27-AUG-1996
DEFINITION DR49 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
H39361
ACCESSION H39361
VERSION H39361.1 GI:915035
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 166)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into PCRscript
Seq primer: KS.

FEATURES

source

Location/Qualifiers

1. .166
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR49"

/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern." 41 a 33 c 32 g 60 t

BASE COUNT

ORIGIN

H39361 Length: 166 February 27, 2002 16:12 Type: N Check: 3789 ..
Found using 'seq2-3' (pappu403.key)

...

18 GTCGGTACAATTCCTCTAATGTACTTCCTCGAATAGATTACTTTTGGTGACGACTTTA
68 73
|-----|

78 TAGAAGATTTCGTACTTCAATACTTTACCGTGTCTCAGTTATCTAAGTAACCTGCTCTATG

138 AATTTCTCATAGCGTCAGTCGTCGTGA
149 154
|-----|

3 matches found in sequence:

h39363 ; TOIG of: h39363 check: 3470 from: 1 to: 157
(from "immun_est.seq")
TOIG of: h39363 check: 3470 from: 1 to: 157

LOCUS H39363 157 bp mRNA EST 27-AUG-1996
DEFINITION DR51 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
H39363
ACCESSION H39363
VERSION H39363.1 GI:915037
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 157)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into PCRscript
Seq primer: KS.

FEATURES

source

Location/Qualifiers

1. .157
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR51"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern." 39 a 43 c 40 g 35 t

BASE COUNT

ORIGIN

H39363 Length: 157 February 27, 2002 16:12 Type: N Check: 3470 ..
Found using 'seq2-3' (pappu403.key)


```
REFERENCE
AUTHORS      Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE        Identification of mRNAs regulated by interferon-gamma in cultured
              rat astrocytes by PCR differential display
JOURNAL      Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE      96437728
COMMENT      Contact: Hart RP
              Rutgers University
              Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
              07102
              Tel: 2016481008
              Fax: 2016481007
              Email: rhart@andromeda.rutgers.edu
              unknown sequence
              subcloned into PCRscript
              Seq primer: KS.
              Location/Qualifiers
                1. .152
                  /organism="Rattus norvegicus"
                  /strain="Sprague-Dawley"
                  /db_xref="taxon:10116"
                  /clone="DR64"
                  /clone_lib="IFNgamma-induced astrocyte ESTs"
                  /notes="cDNA from IFNgamma-induced cultured rat astrocytes
                  was screened by PCR differential display. Regulated cDNAs
                  were amplified, sequenced and confirmed by dot-blot and
                  Northern."
                42 a 38 c 33 g 39 t
              BASE COUNT
              ORIGIN
              H39376 Length: 152 February 27, 2002 16:12 Type: N Check: 8692
              Found using 'seq2-3' (pappu403.key)
              1 GGTAAATGGCTGCTAGTCTGCTACTTCTTACGCACCTTAAGCAAAGTGGGCGATGAAAGA
              |-----|
              49 54
              61 AGCAATCAGCTAGTATTAGTCGACCTCCACGCTCTACTGCAATA
              ...
              -----
              1 match found in sequence:
              h39377 : TOIG of: h39377 check: 8692 from: 1 to: 152
              (from "immun_est.seq")
              TOIG of: h39377 check: 8692 from: 1 to: 152
              LOCUS H39377 152 bp mRNA EST 27-AUG-1996
              DEFINITION DR65 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
              DR65 5'-end with RP similar to unknown sequence, mRNA sequence.
              ACCESSION H39377
              VERSION H39377.1 GI:915051
              KEYWORDS EST.
              SOURCE Norway rat.
              ORGANISM Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
              REFERENCE 1 (bases 1 to 152)
              AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
              TITLE Identification of mRNAs regulated by interferon-gamma in cultured
              rat astrocytes by PCR differential display
              JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
              MEDLINE 96437728
              COMMENT Contact: Hart RP
              Rutgers University
              Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
              07102
              Tel: 2016481008
              Fax: 2016481007
              Email: rhart@andromeda.rutgers.edu
              unknown sequence
```

```
subcloned into PCRscript
Seq primer: KS.
Location/Qualifiers
  1. .152
    /organism="Rattus norvegicus"
    /strain="Sprague-Dawley"
    /db_xref="taxon:10116"
    /clone="DR65"
    /clone_lib="IFNgamma-induced astrocyte ESTs"
    /notes="cDNA from IFNgamma-induced cultured rat astrocytes
    was screened by PCR differential display. Regulated cDNAs
    were amplified, sequenced and confirmed by dot-blot and
    Northern."
  42 a 38 c 33 g 39 t
BASE COUNT
ORIGIN
H39377 Length: 152 February 27, 2002 16:12 Type: N Check: 8692
Found using 'seq2-3' (pappu403.key)
1 GGTAAATGGCTGCTAGTCTGCTACTTCTTACGCACCTTAAGCAAAGTGGGCGATGAAAGA
|-----|
49 54
61 AGCAATCAGCTAGTATTAGTCGACCTCCACGCTCTACTGCAATA
...
-----
2 matches found in sequence:
h39379 : TOIG of: h39379 check: 9105 from: 1 to: 181
(from "immun_est.seq")
TOIG of: h39379 check: 9105 from: 1 to: 181
LOCUS H39379 181 bp mRNA EST 27-AUG-1996
DEFINITION DR67 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR67 5'-end with RP similar to unknown sequence, mRNA sequence.
ACCESSION H39379
VERSION H39379.1 GI:915053
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 181)
AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
JOURNAL Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE 96437728
COMMENT Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into PCRscript
Seq primer: KS.
Location/Qualifiers
  1. .181
    /organism="Rattus norvegicus"
    /strain="Sprague-Dawley"
    /db_xref="taxon:10116"
    /clone="DR67"
    /clone_lib="IFNgamma-induced astrocyte ESTs"
    /notes="cDNA from IFNgamma-induced cultured rat astrocytes
    was screened by PCR differential display. Regulated cDNAs
    were amplified, sequenced and confirmed by dot-blot and
    Northern."
  45 a 44 c 46 g 46 t
BASE COUNT
ORIGIN
```

H39379 Length: 181 February 27, 2002 16:12 Type: N Check: 9105
Found using 'seq2-3' (pappu403.key)

...

64 AGCAACTTTCACACACACGAATTGGCTGCCAACAGCAGGAGGAGAAATCGTCTTTGG
114 119
|-----|
124 GGATGAACCTACGCTCGCTCAACTTTCAGCAACTCTCGGTATTTCTCGTATTTCCC
137 142
|-----|

2 matches found in sequence:

h39381 : TOIG of: h39381 check: 3232 from: 1 to: 170
(from "immun_est.seq")
TOIG of: h39381 check: 3232 from: 1 to: 170

LOCUS H39381 170 bp mRNA EST 27-AUG-1996
DEFINITION DR69 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR69 5'-end with RP similar to unknown sequence, mRNA sequence.

ACCESSION H39381
VERSION H39381.1 GI:915055
KEYWORDS EST.
SOURCE Norway rat.

ORGANISM
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 170)

AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.

TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display

JOURNAL Neuroimmunomodulation 2, 347-355 (1995)

MEDLINE
COMMENT

96437728

CONTACT: Hart RP

Rutgers University

Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ

07102

Tel: 2016481008

Fax: 2016481007

Email: rhart@andromeda.rutgers.edu

unknown sequence

subcloned into PCRscript

Seq primer: KS.

Location/Qualifiers

1. .170

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="DR69"

/clone_lib="IFNgamma-induced astrocyte ESTs"

/notes="cDNA from IFNgamma-induced cultured rat astrocytes

was screened by PCR differential display. Regulated cDNAs

were amplified, sequenced and confirmed by dot-blot and

Northerns."

BASE COUNT 43 a 39 c 45 g 43 t

ORIGIN

H39381 Length: 170 February 27, 2002 16:12 Type: N Check: 322
Found using 'seq2-3' (pappu403.key)

...

67 CATTCTCGCGCACCACACTATTGTGATTCGAGATTCGCTCAGACTGGA
114 119
|-----|
127 GGTATACTATCGTACCGCGGTATACGGGTGAGGTGGCTTAACGTTT
164 169
|-----|

2 matches found in sequence:

h39383 : TOIG of: h39383 check: 221 from: 1 to: 174
(from "immun_est.seq")
TOIG of: h39383 check: 221 from: 1 to: 174

LOCUS H39383 174 bp mRNA EST 27-AUG-1996
DEFINITION DR71 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR71 5'-end with RP similar to unknown sequence, mRNA sequence.

ACCESSION H39383

VERSION H39383.1 GI:915057

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 174)

AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.

TITLE Identification of mRNAs regulated by interferon-gamma in cultured

rat astrocytes by PCR differential display

JOURNAL Neuroimmunomodulation 2, 347-355 (1995)

MEDLINE
COMMENT

96437728

CONTACT: Hart RP

Rutgers University

Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ

07102

Tel: 2016481008

Fax: 2016481007

Email: rhart@andromeda.rutgers.edu

unknown sequence

subcloned into PCRscript

Seq primer: KS.

Location/Qualifiers

1. .174

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="DR71"

/clone_lib="IFNgamma-induced astrocyte ESTs"

/notes="cDNA from IFNgamma-induced cultured rat astrocytes

was screened by PCR differential display. Regulated cDNAs

were amplified, sequenced and confirmed by dot-blot and

Northerns."

BASE COUNT 43 a 41 c 44 g 46 t

ORIGIN

H39383 Length: 174 February 27, 2002 16:12 Type: N Check: 221

Found using 'seq2-3' (pappu403.key)

...

58 AGCAACTTTCACACACACGAATTGGCTGCCAACAGCAGGAGGAGAAATCGTCTTTGG
108 113
|-----|

118 GGATGAACCTACGCTCGCTCAACTTTCAGCAACTCTCGGTATTTCTCGTATTTCC
131 136
|-----|

1 match found in sequence:

h39388 : TOIG of: h39388 check: 6165 from: 1 to: 171
(from "immun_est.seq")
TOIG of: h39388 check: 6165 from: 1 to: 171

LOCUS H39388 171 bp mRNA EST 27-AUG-1996

DEFINITION DR78 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone

DR78 5'-end with RP similar to ATP synthase subunit e, mRNA

sequence.

ACCESSION H39388

VERSION H39388.1 GI:915062

KEYWORDS
SOURCE
ORGANISM

EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 171)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728

Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
high similarity to rat mRNA for ATP synthase subunit e
subcloned into pCRscript
Seq primer: KS.

FEATURES
source

Location/Qualifiers
1. .171

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR78"

/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern analysis."

BASE COUNT 56 a 20 c 56 g 39 t
ORIGIN

H39388 Length: 171 February 27, 2002 16:12 Type: N Check: 6165 ..
Found using 'seq2-3' (pappu403.key)

...

87 GAAGATGTCAGCATTTTCAAGTAGGCGGCGAGTGATCTGTTCTCTCTGTTGAGG
137 142
|-----|

147 ATGAATAAGCTTCATTGTTGNA

1 match found in sequence:

h39390 ; TOIG of: h39390 check: 5722 from: 1 to: 155
(from "immun_est.seq")

TOIG of: h39390 check: 5722 from: 1 to: 155

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

H39390 155 bp mRNA EST 27-AUG-1996
DR80 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR80 5'-end with RP similar to unknown sequence, mRNA sequence.

Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into pCRscript
Seq primer: KS.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 155)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728

Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ

07102

Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into pCRscript
Seq primer: KS.

FEATURES
source

Location/Qualifiers
1. .155

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR80"

/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern analysis."

BASE COUNT 49 a 43 c 30 g 33 t
ORIGIN

H39390 Length: 155 February 27, 2002 16:12 Type: N Check: 5722 ..
Found using 'seq2-3' (pappu403.key)

1 TAGCAAAAGATATTATACCTGTGTATATATGTCATCAGGGCGGACATATATACACACTA
41 46
|-----|

61 AAACAAGGCTGGTGACAAACAGACTGCACACCACAG

...

2 matches found in sequence:
h39393 ; TOIG of: h39393 check: 3232 from: 1 to: 170
(from "immun_est.seq")
TOIG of: h39393 check: 3232 from: 1 to: 170

LOCUS H39393 170 bp mRNA EST 27-AUG-1996

DEFINITION DR83 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR83 5'-end with RP similar to unknown sequence, mRNA sequence.

ACCESSION H39393

VERSION H39393.1 GI:915067

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 170)

AUTHORS Kuchinke,W., Hart,R.P. and Jonakait,G.M.

TITLE Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display

JOURNAL Neuroimmunomodulation 2, 347-355 (1995)

MEDLINE 96437728

COMMENT

Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102

Tel: 2016481008

Fax: 2016481007

Email: rhart@andromeda.rutgers.edu

unknown sequence

subcloned into pCRscript

Seq primer: KS.

FEATURES
source

Location/Qualifiers
1. .170

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="DR83"

/note="cDNA from IFNgamma-induced astrocyte ESTs"
/clone_lib="IFNgamma-induced cultured rat astrocytes"

was screened by PCR differential display. Regulated cDNAs were amplified, sequenced and confirmed by dot-blot and Northern's.

BASE COUNT 43 a 39 c 45 g 43 t
ORIGIN

H39393 Length: 170 February 27, 2002 16:12 Type: N Check: 3232 ..
Found using 'seq2-3' (pappu403.key)

...

67 CATTGTCGCGCACCACACTATTGTTGATTCGAGATTCCTGCACAGACTGGAGCCCTTAGA
117 122 |-----|

127 GGTATACATCGTACGCCGTATACGGCTGAGGTGGCTAAGCTTT
164 169 |-----|

2 matches found in sequence:

h39395 ; TOIG of: h39395 check: 221 from: 1 to: 174
(from "immun_est.seq")

TOIG of: h39395 check: 221 from: 1 to: 174

LOCUS H39395 174 bp mRNA EST 27-AUG-1996
DEFINITION DR85 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR85 5'-end with RP similar to unknown sequence, mRNA sequence.

ACCESSION H39395

VERSION H39395.1 GI:915069

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 174)

Kuchinke,W., Hart,R.P. and Jonakait,G.M.

Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display

Neuroimmunomodulation 2, 347-355 (1995)

96437728

COMMENT

Contact: Hart RP

Rutgers University

Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ

07102

Tel: 2016481008

Fax: 2016481007

Email: rhart@andromeda.rutgers.edu

unknown sequence

subcloned into PCRscript

Seq primer: KS.

FEATURES

source

1. 174

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="DR85"

/note="CDNA from IFNgamma-induced astrocyte ESTs"

was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern's.

BASE COUNT 43 a 41 c 44 g 46 t

ORIGIN

H39395 Length: 174 February 27, 2002 16:12 Type: N Check: 221 ..
Found using 'seq2-3' (pappu403.key)

58

ACCAACTTTCACACACAGCAATTGGCTGCCCAACAGGCAGGAGGAGAAATCGTCTTGG
|-----|

108 113

118 GGATGAAACTACCGTCGCTCAACTTTGAGCAACTCTCGTATTCTCGTGATTTC
131 136 |-----|

1 match found in sequence:

h39405 ; TOIG of: h39405 check: 4631 from: 1 to: 159
(from "immun_est.seq")

TOIG of: h39405 check: 4631 from: 1 to: 159

LOCUS H39405 159 bp mRNA EST 27-AUG-1996
DEFINITION DR97 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR97 5'-end with RP similar to unknown sequence, mRNA sequence.

ACCESSION H39405

VERSION H39405.1 GI:915079

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 159)

Kuchinke,W., Hart,R.P. and Jonakait,G.M.

Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display

Neuroimmunomodulation 2, 347-355 (1995)

96437728

COMMENT

Contact: Hart RP

Rutgers University

Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ

07102

Tel: 2016481008

Fax: 2016481007

Email: rhart@andromeda.rutgers.edu

unknown sequence

subcloned into PCRscript

Seq primer: KS.

FEATURES

source

1. 159

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="DR97"

/note="CDNA from IFNgamma-induced astrocyte ESTs"

was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern's.

BASE COUNT 54 a 28 c 54 g 23 t

ORIGIN

H39405 Length: 159 February 27, 2002 16:12 Type: N Check: 4631 ..
Found using 'seq2-3' (pappu403.key)

1 GAGCTGGGGGATGGGGTAACATGGGGTTGTAAACACCGTGAGGACGACGCTGCTCAGA
|-----|
47 52

61 GCAAAATAGGAGGCATTAAAGCAAAACCCACACATTCTGAA

...

1 match found in sequence:

h39412 ; TOIG of: h39412 check: 8928 from: 1 to: 149
(from "immun_est.seq")

TOIG of: h39412 check: 8928 from: 1 to: 149

LOCUS H39412 149 bp mRNA EST 27-AUG-1996

DEFINITION DR104 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone

DR104 5'-end similar to unknown sequence, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

H39412
H39412.1 GI:915086
EST.
Norway rat.
Rattus norvegicus

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 149)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
direct sequencing
Seq primer: CAGCTGCTTG.
Location/Qualifiers
1. .149
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR104"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northernblots."
Northernblots."
Northernblots."

FEATURES
source

Location/Qualifiers
1. .149
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR104"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northernblots."
Northernblots."
Northernblots."

BASE COUNT
ORIGIN

26 a 34 c 42 g 47 t
H39412 Length: 149 February 27, 2002 16:12 Type: N Check: 8928 ..
Found using 'seq2-3' (pappu403.key)
...

40 CTCAGTCAGGGTAGGATGGGACCACATCTTTCTAGTGAAGACTCATGGGCGCCTCA
90 95
|-----|
|-----|

100 TTTTAAAGTTTGACATCTCCCTGGGAGCTAAGGCTTGGCATCG

BASE COUNT
ORIGIN

145 bp mRNA EST 27-AUG-1996
H39415
DR109 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR109 5'-end similar to unknown sequence, mRNA sequence.
H39415
ACCESSION H39415.1 GI:915090
VERSION H39415.1
KEYWORDS EST.
SOURCE Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 145)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)

MEDLINE
COMMENT

96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
subcloned into PCRscript
Seq primer: -40 primer.
Location/Qualifiers
1. .145
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR109"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northernblots."
Northernblots."
Northernblots."

FEATURES
source

Location/Qualifiers
1. .145
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DR109"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northernblots."
Northernblots."
Northernblots."

BASE COUNT 25 a 32 c 40 g 48 t
ORIGIN

H39415 Length: 145 February 27, 2002 16:12 Type: N Check: 9554 ..
Found using 'seq2-3' (pappu403.key)
...

41 CTCAGTCAGGGTAGGATGGGACCACATCTTTCTAGTGAAGACTCATGGGCGCCTCA
91 96
|-----|
|-----|

101 TTTTAAAGTTTGACATCTCCCTGGGAGCTAAGGCTTGGCTCTT

BASE COUNT
ORIGIN

74 bp mRNA EST 27-AUG-1996
H39417
DR111 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR111 5'-end with RP similar to unknown sequence, mRNA sequence.
H39417
ACCESSION H39417.1 GI:915092
VERSION H39417.1
KEYWORDS EST.
SOURCE Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 74)
Kuchinke,W., Hart,R.P. and Jonakait,G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: rhart@andromeda.rutgers.edu
unknown sequence
direct sequencing
Seq primer: TCCACAGGTC.
Location/Qualifiers
1. .74
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"

FEATURES
source

Location/Qualifiers
1. .74
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"

```
/db_xref="taxon:10116"
/clone="DR111"
/clone_lib="IFNgamma-induced astrocyte ESTs"
/note="cDNA from IFNgamma-induced cultured rat astrocytes
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern."
BASE COUNT      16 a   13 c   22 g   23 t
ORIGIN
H39417 Length: 74 February 27, 2002 16:12 Type: N Check: 967
Found using 'seq2-3' (pappu403.key)

1 TCATTCCGGAGGCTGTAGGTAGGTAGTTTAGGAGATTGGCGACTCCCGAAGTGATT
41 46
61 CATATGCTTCAGAC

-----
1 match found in sequence:
h39419 ; TOIG of: h39419 check: 8631 from: 1 to: 129
(from "immun_est.seq")
TOIG of: h39419 check: 8631 from: 1 to: 129

LOCUS      H39419      129 bp      mRNA      EST      27-AUG-1996
DEFINITION Dr113 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
ACCESSION H39419
VERSION   H39419.1 GI:915094
KEYWORDS  EST.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 129)
AUTHORS   Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE     Identification of mRNAs regulated by interferon-gamma in cultured
           rat astrocytes by PCR differential display
JOURNAL   Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE   9643728
COMMENT   Contact: Hart RP
           Rutgers University
           Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
           07102
           Tel: 2016481008
           Fax: 2016481007
           Email: rhart@andromeda.rutgers.edu
           unknown sequence
           direct sequencing
           Seq primer: TCCACAGGTC.
           Location/Qualifiers
             1. .129
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="DR113"
               /clone_lib="IFNgamma-induced astrocyte ESTs"
               /note="cDNA from IFNgamma-induced cultured rat astrocytes
               was screened by PCR differential display. Regulated cDNAs
               were amplified, sequenced and confirmed by dot-blot and
               Northern."
BASE COUNT      25 a   32 c   39 g   33 t
ORIGIN
H39419 Length: 129 February 27, 2002 16:12 Type: N Check: 8631
Found using 'seq2-3' (pappu403.key)

...

10 GCTCGGGGCTGGTCTTGTTCTCGATGGGAGGAGATTTCGGGATCAAGCGATGCC
|-----|
```

```
70 TGATCTGTATCCTCACCTCCACCCTCCAGATGCAGTCTTGCC
...
-----
1 match found in sequence:
h39424 ; TOIG of: h39424 check: 3374 from: 1 to: 126
(from "immun_est.seq")
TOIG of: h39424 check: 3374 from: 1 to: 126

LOCUS      H39424      126 bp      mRNA      EST      27-AUG-1996
DEFINITION Dr118 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
ACCESSION H39424
VERSION   H39424.1 GI:915099
KEYWORDS  EST.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 126)
AUTHORS   Kuchinke,W., Hart,R.P. and Jonakait,G.M.
TITLE     Identification of mRNAs regulated by interferon-gamma in cultured
           rat astrocytes by PCR differential display
JOURNAL   Neuroimmunomodulation 2, 347-355 (1995)
MEDLINE   9643728
COMMENT   Contact: Hart RP
           Rutgers University
           Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
           07102
           Tel: 2016481008
           Fax: 2016481007
           Email: rhart@andromeda.rutgers.edu
           unknown sequence
           direct sequencing
           Seq primer: AGCGCTGGAG.
           Location/Qualifiers
             1. .126
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="DR118"
               /clone_lib="IFNgamma-induced astrocyte ESTs"
               /note="cDNA from IFNgamma-induced cultured rat astrocytes
               was screened by PCR differential display. Regulated cDNAs
               were amplified, sequenced and confirmed by dot-blot and
               Northern."
BASE COUNT      28 a   37 c   24 g   37 t
ORIGIN
H39424 Length: 126 February 27, 2002 16:12 Type: N Check: 3374
Found using 'seq2-3' (pappu403.key)

1 CGACGGCTAGTTTCATCCCAAGACGATTTCTCTCGCTTTGTTGGCAGCAATTCGTT
23 28
61 TGTTGTGAATTCCTTCT
...
-----
1 match found in sequence:
h39426 ; TOIG of: h39426 check: 8706 from: 1 to: 180
(from "immun_est.seq")
TOIG of: h39426 check: 8706 from: 1 to: 180

LOCUS      H39426      180 bp      mRNA      EST      27-AUG-1996
DEFINITION DR93 IFNgamma-induced astrocyte ESTs Rattus norvegicus cDNA clone
DR93 5'-end with RP, mRNA sequence.
```